

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

STAYING ALIVE

The Mexican cavefish that defies diabetes to thrive with high blood sugar **PAGES 595 & 647**



PREHISTORY

DIVIDED BY DNA

Archaeologists wrestle with power of ancient genomes

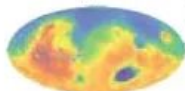
PAGE 573

PLANETARY SCIENCE

MARTIAN CHRONICLE

An early date for the formation of oceans on Mars

PAGES 590 & 643



NEUROSCIENCE

MOBILE BRAIN SCANS

Wearable system allows imaging on the move

PAGE 657

NATURE.COM/NATURE

29 March 2018

Vol. 555, No. 7698

THIS WEEK

EDITORIALS

AL FUNGHI Beef and mushroom blend is latest superfood **p.560**

WORLD VIEW Rein in the profiteers in the drug industry **p.561**



TAKE AWAY Coloured powder reveals how far bats fly for food **p.563**

Use and abuse of ancient DNA

Researchers in several complementary disciplines need to tread carefully over the shared landscapes of the past.

History might, as historian Arnold Toynbee allegedly said, be one damned thing after another, but historians and archaeologists spend a lot of their time trying to put those things into the right order. Assistance from science over the decades has been transformative, but not without difficulty: it took years for some archaeologists to be won over by radiocarbon dating.

Now, historians and archaeologists are grappling with a new scientific technique. As we discuss in a News Feature on page 573, the genetic study of ancient DNA is exploding, and the findings are posing several problems. One is a need for geneticists, archaeologists, historians and anthropologists to understand exactly how their skills and insights complement each other's. It is clear, for example, that although genetics has useful things to say about the sweep of population history, the more conventional disciplines provide essential context.

Another problem is fear that simplistic takes on ancient DNA will mirror damaging uses of the idea of 'culture history'. Culture history views the discovery of old artefacts as a proxy for the movement of the people who made them. According to this idea, a particular floral design on a pot that spread from south to north over a few centuries, for example, would indicate that the specific group of people that painted it was on the move — and carried the design with it.

These fears are not just about scholarship. Simplistic readings of culture history have encouraged people with political agendas to falsely draw clear boundaries between the behaviour and the claimed territory of some ancient (and not-so-ancient) populations — and to infer similarities with their claimed modern equivalents. For example, they often refer to the work of early-twentieth-century German archaeologist Gustaf Kossinna, who used culture history to trace the supposed origins of modern Germany to the spread of Corded Ware, a type of ceramic found throughout central Europe in the Bronze Age. Kossinna's ideas, although influential, have proved to be scientifically simplistic. They became notorious following their use by the Nazi party to legitimize its territorial goals and beliefs about the racial superiority of German-speaking peoples.

Scholars are anxious because extremists are scrutinizing the results of ancient-DNA studies and trying to use them for similar misleading ends. Ancient DNA, for example, offers evidence of large migrations that coincide with cultural changes in the archaeological record, including the emergence of Corded Ware. Some archaeologists have expressed fears that the extremists will wrongly present such conclusions as backing for Kossinna's theories.

Another problem for archaeologists and historians relates to the potential for abuse of the results of ancient-DNA studies looking at more recent times, such as the Migration Period around the fall of the Roman Empire or the era covered by the Viking sagas. They worry that DNA studies of groups described as Franks or Anglo-Saxons or Vikings will reify them by attaching misleading genetic profiles to categories that were devised by historians, and are not representative of how individuals viewed

themselves at the time. Already, some people have picked up on such studies as a way to try to trace their roots to such supposed populations, to justify claims they have a right to some territory or other (L.-J. Richardson and T. Booth *Papers Inst. Archaeol.* 27, 25; 2017).

On the contrary, genetic and historical evidence suggests that there was widespread mixing during these periods, across populations and geography. Indeed, presented correctly alongside insights from other disciplines, ancient-DNA research can be a powerful weapon against bigotry. Studies documenting migrations can drive home the point that present-day peoples in one area often share few genetic links with ancient peoples who lived in the same place. And when they do focus on relatively recent times, DNA projects can highlight the diversity of past peoples who otherwise might be seen as

homogenous. A 2016 study of Anglo-Saxon burials, for example, found a mix of ancestry, with some people related to earlier inhabitants of England and others tracing their ancestry across the Channel (S. Schiffels *et al. Nature Commun.* 7, 10408; 2016).

Two recommendations can be made for the public behaviour of scientists and other scholars. The first: give ample credit to the insights of complementary disciplines. The second: refute statements that misconstrue what your insights actually reveal and that can be used politically to justify disrespect, or worse, to groups of people. ■

Digital trust

A scandal over an academic's use of Facebook data highlights the need for research scrutiny.

Revelations keep emerging in the Cambridge Analytica personal-data scandal, which has captured global public attention for more than a week. But when the dust settles, researchers harvesting data online will face greater scrutiny. And so they should.

At the centre of the controversy is Aleksandr Kogan, a psychologist and neuroscientist at the University of Cambridge, UK. In 2014, he recruited people to complete a number of surveys and sign up to an app that handed over Facebook information on themselves — and tens of millions of Facebook friends. Kogan passed the data to SCL, a UK firm that later founded controversial political-consultancy firm Cambridge Analytica in London. (All those involved deny any wrongdoing.)

Last week, Facebook announced restrictions on data harvesting by third parties, including drastically reducing the kinds of information

that app developers can access. (It had already changed its rules in 2014 to stop developers gleaning data from users' friends through their apps.) But damage has been done: the public has good reason to be angry about the way in which researchers and companies have seemingly used personal data without consumers' full understanding or consent.

Where do academic researchers fit in? Handled correctly, online data can be a major boon to research, and the world would benefit from companies such as Facebook making their data more open. Ethical safeguards for research that intervenes in human lives were largely set up for medical and psychological studies, and are often written with definitions that exclude Internet research. In the United States, for example, unless data collected are both private and identifiable, informed consent is usually not deemed necessary, and research requires minimal, if any, oversight by an institutional review board. This would include data from Twitter, which are by default public. Models built on anonymized Facebook data would also tend to be exempt.

Kogan's study was unusual, both in that it was done by a university academic for a private company he operated, and in that the data were passed to a third party. Yet there is a common theme behind this controversy and ones that preceded it — such as a study warning that someone's sexual orientation could be determined from their online presence (Y. Wang and M. Kosinski *J. Personality Soc. Psychol.* **114**, 246–257; 2018). Data were used in ways well beyond what users expected or intended. Bundled together and trawled by algorithms, innocuous data points can reveal information that users might reasonably expect to stay private and that might be used in ways they are not happy with.

Guidance does exist. A number of projects are grappling with the ethical challenges of big data. US and European funders have supported efforts in this area, and have issued recommendations such as rethinking what counts as 'public' data and the need to consider a study's potential harm to society, as well as to individuals. (The University of Cambridge is among the institutions writing guidelines for Internet-mediated research, after the UK Research Integrity Office issued non-binding recommendations on the topic in 2016.) Funders should further support such efforts, and make them better known to researchers.

“Innocuous data points can reveal personal information.”

Sticking points remain, a major one being that consent is often not practical when retrospectively accessing data from millions of individuals. But as outlined for biomedical scientists in the 1978 Belmont Report, the principle of beneficence applies: researchers should put the good of research participants first and, with that in mind, perform their own assessment of risks versus benefits. Studies should not be done just because the data are there. In studies that are too large to ask participants for consent, researchers should poll the views of samples of subjects and of any population that could be affected by the outcomes. Ethics training on research should be extended to computer scientists who have not conventionally worked with human study participants.

Academics across many fields know well how technology can outpace its regulation. All researchers have a duty to consider the ethics of their work beyond the strict limits of law or today's regulations. If they don't, they will face serious and continued loss of public trust. ■

Burger al funghi

Mushroom–beef blends can tackle expanding waistlines and carbon footprints.

Big burgers and petrol-guzzling vehicles: it's a match made in the United States. But last month, the US fast-food chain Sonic Drive-In tweaked its menu to introduce something different: a burger with some of its beef replaced with mushrooms. The Signature Slinger cheeseburger, Sonic promises, will deliver drive-through customers “all of the flavor with none of the guilt”.

The claimed conscience-easing credentials of the Slinger are twofold: reduced calories and a smaller carbon footprint. It's targeted at those who wouldn't switch to a non-meat alternative, such as soya. And the burger does not come alone. It's merely the latest offering from a movement that has united chefs, scientists and health and environmental advocates. Led by the Culinary Institute of America in Hyde Park, New York, with predictable support from mushroom producers, the initiative seeks to promote both public health and environmental sustainability, by replacing 100% ground beef in some foods with a blend that contains up to half mushrooms. It's the biggest new foodie trend that you haven't heard of. Harvard University in Cambridge, Massachusetts, and the University of Southern California in Los Angeles are among the institutions that have started to cook with it routinely.

Does the science support the claims?

Beef has an outside environmental footprint. Per gram of protein, beef requires on average 50 times more land and produces 100 times more greenhouse gases than do beans and other plant-based proteins. Replacing 30% of the beef with mushrooms in the roughly 10 billion burgers that Americans eat each year would reduce emissions equivalent to taking 2 million cars off the road, according to the World Resources Institute, an environmental think tank in Washington DC.

Still, mushroom cultivation is not a green panacea. Mushrooms are fussy about temperature and are usually grown inside energy-hungry climate-controlled sheds. A 2012 analysis reported that mushrooms

have a greenhouse-gas impact almost ten times that of vegetables such as onions, carrots and cabbages (M. Berners-Lee *et al. Energy Policy* **43**, 184–190; 2012).

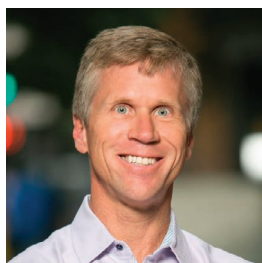
Although individual consumers who choose a mushroom substitute might feel less guilty, demand for beef continues to rise alongside incomes around the world. As such, scientists and governments must continue to look for ways to reduce the impact of beef farming itself. Expanding cattle pasture is the main driver of deforestation in the Brazilian Amazon, and one-third of the world's cropland goes to feeding animals — usually cattle. Research is needed to raise pasture productivity, improve feed and reduce methane emissions. So, too, is a renewed effort to combat deforestation.

On the health front, research has shown that substituting one-quarter of the beef in meals with mushrooms decreases caloric intake by about one-third. Saturated fats also drop, and chefs can often make do with less salt. Tests across US school dinners suggest that these benefits seem to come without a negative impact, such as leaving pupils hungry.

In an era of expanding waistlines, this is potentially good news. Earlier this month, health authorities in the United Kingdom called on food manufacturers to voluntarily reduce the calories in processed products by 20% by 2024. Doing so would save nearly £9 billion (US\$12.7 billion) and prevent more than 35,000 premature deaths due to obesity-related illness over 25 years, Public Health England estimated. Sonic's blended burger is substantially smaller than its regular cheeseburger and would more than deliver on that goal, starting at around 340 calories compared with 580 calories for the chain's standard.

Will it and other mushroom-blended meats succeed where health and environmental advocates have failed, and persuade people to eat less beef? Certainly, it seems a better approach than simply asking people to change their ways by pointing out the cost.

But, as with zero-alcohol wine and sugar-free soft drinks, taste will ultimately determine consumer choice. So how does the Signature Slinger score? In a blind taste comparison, two *Nature* reporters were both able to correctly identify the mushroom blend. Setting aside the fact that both patties had the distinct taste and texture of fast food, one person preferred the pure beefburger, and the other the mushroom blend. Both agreed that if the blended burger was the future, then the future was better with (guilt-inducing) bacon. ■



Drug executives should take a Hippocratic oath

The industry must earn patients' trust that new medicines really are worth the price, says Bob More.

This month, pharmaceutical executive Martin Shkreli was sentenced to seven years in prison for securities fraud. But those charges are not what gained him notoriety. In August 2015, his company bought monopoly rights to a life-saving drug that treats parasitic infections in people with weak immune systems. Without adding any value or doing significant research, the company then raised the drug's price by more than 5,000%, from US\$13.50 to \$750 for a single tablet.

Plenty of drug-industry insiders decried Shkreli as “not one of us”, emphasizing the virtues of biotechnology and innovative pharmaceuticals. I am a fan of the drug industry — I invest in promising leads and start-ups. But the public sees our industry as one entity, not an industry of good and bad players. When any executive engages in price gouging, it is a reflection on us all, and validates the negative bias against our industry. Innovators and opportunists will go down together, unless the former agree to push the latter out.

Yes, the drug industry is a business. But it is a business with a beautiful purpose: treating what ails people. As a venture capitalist, I look for ideas that increase value. If we succeed in finding new and better therapies, we should be rewarded for the risks and investments we make. A course of the hepatitis-C drug sofosbuvir (Sovaldi) can run, controversially, into the tens of thousands of dollars — but the drug took years of cutting-edge research to produce and cures a disease that otherwise requires a liver transplant, which costs upwards of \$500,000.

As George Merck, president of drug company Merck from 1925 to 1950, put it: “We try never to forget that medicine is for the people. It is not for the profits. The profits follow, and if we have remembered that, they have never failed to appear. The better we have remembered that, the larger they have been.”

Passed down from Merck to executives such as Roy Vagelos, a former head of Merck, and George Yancopoulos, now chief scientific officer at Regeneron Pharmaceuticals, is a fierce belief that if you do the science right, you will ultimately deliver innovative drugs to patients that will save lives and cure disease. I feel confident that investments (some by my firm) in science-driven companies such as Genentech, Kite Pharma and Achaogen will produce results. So will a new generation of companies, such as eGenesis, Encoded Genomics and Vir Biotechnology, which work, respectively, on producing transplant organs, ways of regulating genes and fighting infection. It takes people with their kind of scientific fanaticism and high standards to hold our collective feet to the fire.

I think there are many more George Mercks in the drug industry than Martin Shkrelis. But too many in the industry, driven by the demands of Wall Street, have shifted to a model that seeks to maximize profit over helping patients. Among others, Michael Pearson, former chief

executive at Valeant Pharmaceuticals, has been publicly criticized for jacking up prices on several drugs. And I have, unfortunately, been privy to conversations in which executives planned price rises to force wholesalers to make larger purchases so they can hit sales expectations.

None of us wants to believe that we are part of the problem. But as long as we tolerate the presence of these bad actors, we are all complicit.

An academic colleague once asked me about violating a company's patents because, he assumed, “you guys in business do that all the time”. No, we don't. We depend on patents. Responsibility in our business means following the laws that safeguard our own intellectual property when the time comes.

Over the long term, mistrust raises the cost of capital and makes it difficult to fund worthy science. Just two weeks ago, the US govern-

ment charged Theranos and its chief executive, Elizabeth Holmes, with defrauding investors. (She and the company reached a settlement without denying or admitting wrongdoing.)

In my experience, the biggest drag on our stock prices is our bouts of dishonesty with Wall Street, physicians and patients. And, ultimately, it's the patients who suffer.

The innovations ahead of us are both breathtakingly effective and expensive: gene editing, chimaeric organs, personalized immunotherapy. But members of the public will not support this research if they suspect that they are being gamed. Transparency about pricing decisions, while sometimes painful in the short run, builds trust.

So, what to do? When we see practices that put profits ahead of patient needs, we need to call foul publicly and not look the other way. We need to

move away from gimmicks that raise only the price of drugs, not their value. People who use these tricks should be fired. Their projects should be shunned by venture capitalists and potential funding partners. We should lobby for a patent system that rewards real novelty, not more years of monopoly for a mint-flavoured version of an existing drug. (I exaggerate, but many of the reformulations used to retain rights are incremental improvements, at best.) Trade groups such as the Biotechnology Industry Organization and PhRMA, both in Washington DC, should emphasize this in their codes of ethics, and members should hold each other to account.

If the pharmaceutical industry is to thrive, we must all embrace the Hippocratic oath, whether in academia, business, finance or at the bench. The cost of lost trust to the drug industry is worth much more than manipulated short-term gains. ■

Bob More is a managing partner at venture-capital firm Alta Partners in San Francisco, California.
e-mail: bmore@altapartners.com

WHEN WE SEE
PRACTICES THAT PUT
PROFITS
AHEAD OF
PATIENTS,
WE NEED TO
CALL FOUL
PUBLICLY.

SEVEN DAYS

The news in brief

SPACE

Exoplanet probe

The European Space Agency will launch the world's first space telescope dedicated to studying the atmospheres of exoplanets, after it selected the mission for development on 20 March. The mission, called Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL), will probe more than 1,000 known exoplanet systems. See page 571 for more.

HEALTH

Yellow-fever rise

In response to an unusually severe outbreak of yellow fever, Brazil's health minister Ricardo Barros announced on 20 March that the government intends to vaccinate the country's entire population against the disease. Since July 2017, the virus has infected 920 people and killed 300. Officials recorded 196 deaths over the same period in the previous year. The decision to expand yellow-fever vaccinations was not made sooner because it is not without risk: one recipient in 100,000 develops dangerous reactions, such as hepatitis. The government hopes that, by April 2019, it will have reached the 78 million people who remain unvaccinated.

Guinea worm

South Sudan has ended the spread of Guinea-worm disease within its borders, leaving Chad and Ethiopia as the only two countries with ongoing spread of the debilitating parasitic infection. South Sudan has gone 15 consecutive months — longer than the worm's life cycle — without recording a case, said the Carter Center in Atlanta, Georgia, on 21 March; the organization is leading a global effort to



MELANIE STETSON FREEMAN/THE CHRISTIAN SCIENCE MONITOR VIA GETTY

Weather costs for 2017 break records

The costs of severe weather and climate events hit a record high of US\$320 billion in 2017, the World Meteorological Association (WMO) said in its annual state-of-the-climate report on 22 March. The bulk of the bill — some \$265 billion — was attributable to damages in the United States from hurricanes Harvey, Irma and Maria. The WMO also cited the costs

of monsoon flooding in Asia and drought in eastern Africa (pictured, the dry Mandrare River in Madagascar). Overall, 2017 was the third-hottest year on record, with global average temperatures at roughly 1.1 °C above the pre-industrial average, according to the report. The nine warmest years on record have all occurred since 2005.

eradicate Guinea worm. The World Health Organization in Geneva, Switzerland, must still certify that the disease has been eradicated in South Sudan, which involves a much longer process. Kenya was deemed free of Guinea worm only in February, despite last recording a case in 1994.

CLIMATE

Frozen records

Arctic sea ice reached its peak winter extent on 17 March, marking the second-lowest maximum in the 39-year satellite record. At 14.48 million square kilometres, this year's peak is just 0.4% above last year's record-low maximum extent. The four lowest maxima all

happened in the past four years, the US National Snow and Ice Data Center said on 23 March. This winter saw parts of the Arctic freezing up relatively late, followed by warm air temperatures throughout much of the season.

EVENTS

Brexit deal urged

UK politicians have urged the government to make an early deal with the European Union on science. Members of Parliament on the House of Commons Science and Technology Committee warned that post-Brexit arrangements for research and innovation could become a bargaining chip in wider trade negotiations unless

research talks are made an urgent priority. Leaving a deal on science until later in the year risks damaging the United Kingdom's status "as a science superpower", they argued in a report published on 19 March. Leading concerns include clarifying whether the government will seek to join the European Commission's next research-funding programme, and whether there will be special immigration arrangements for scientists.

Whale stranding

Roughly 150 short-finned pilot whales (*Globicephala macrorhynchus*) were found stranded on an Australian beach 315 kilometres south of Perth on the morning of 23 March. Animal workers

VICAR/ALAMY

and volunteers were able to rescue and release just seven whales back to deeper waters by nightfall; the rest died. Three whales that had been released returned to the beach the following day; they were humanely killed. All but two of the dead whales have been removed from the beach at Hamelin Bay, near the continent's southwestern tip. The area was also the site of a stranding of 87 pilot whales and dolphins in 2009.

UNIVERSITIES

Science campus

The French government has endorsed an international panel's recommendation to delay for 30 months a decision on whether to give permanent 'excellence' status to the Paris-Saclay University cluster, part of the Saclay science super-campus southwest of Paris (**pictured**). The newly established cluster, a grouping of 14 institutions, has temporary excellence status under the government's IDEX initiative and receives €30 million (US\$37 million) in funding a year; permanent status would guarantee that income. The cluster has been on probation since 2012, and must become an integrated establishment by the 2020 academic year, the panel said. The Saclay science project initially aimed to bring some



20 research and teaching institutions under one umbrella. But progress was stalled by reluctance from elite institutions known as *grandes écoles* to give up their names and autonomy. Last October, President Emmanuel Macron announced that the project would proceed with two clusters: one called Paris-Saclay University, the other called NewUni.

FUNDING

US budget deal

Nearly all US science agencies would see their budgets grow in 2018, under a US\$1.3-trillion deal that President Donald Trump signed into law on 23 March. The legislation, which funds government operations up to 30 September, would boost funding for the National Institutes of Health to \$37 billion, \$3 billion above the 2017 level. The National Science Foundation would get

\$7.8 billion, up \$295 million, and NASA's budget would rise by \$1.1 billion, to \$20.7 billion. For the second year running, lawmakers in Congress ignored the steep cuts to science agencies sought by Trump. See page 572.

ENVIRONMENT

Declining diversity

Human activities are threatening the security of food and clean water by causing a rapid decline in the variety of plants and animals on Earth, according to United Nations-backed assessments released on 23 March. Conducted by more than 500 experts across 100 countries, the 3-year study looked at all areas of the world except the poles and the open ocean. Researchers found that in nearly every region, pollution, climate change and unsustainable exploitation of natural resources are endangering biodiversity. The

assessments estimate that if the situation continues as it is, by 2050 there will be no exploitable fish stocks in the Asia-Pacific region, and the Americas will lose 40% of their biodiversity.

PEOPLE

Public-health chief

HIV researcher Robert Redfield will lead the US Centers for Disease Control and Prevention (CDC), the secretary of health and human services, Alex Azar, announced on 21 March. Redfield "has dedicated his entire life to promoting public health and providing compassionate care to his patients", Azar said in a statement. Redfield, co-director of the Institute of Human Virology at the University of Maryland in Baltimore, has studied HIV epidemiology, pathology and vaccines for over 30 years. He replaces Brenda Fitzgerald, who resigned on 31 January following news reports that she had traded stock in tobacco companies while leading the agency. See go.nature.com/2pbabvn for more.

FACILITIES

Sports brain bank

Australia's first bank for collecting the brains of sportspeople opened on 27 March. The Australian Sports Brain Bank will enable research into chronic traumatic encephalopathy. The condition affects people with a history of repeated brain injury, such as those who have played contact sports, and military veterans. The bank, a joint initiative of the University of Sydney's Brain and Mind Centre and the city's Royal Prince Alfred Hospital, has so far received donation pledges from six Australian former sports professionals. It will also collect brains from people who haven't played sport professionally.

► NATURE.COM

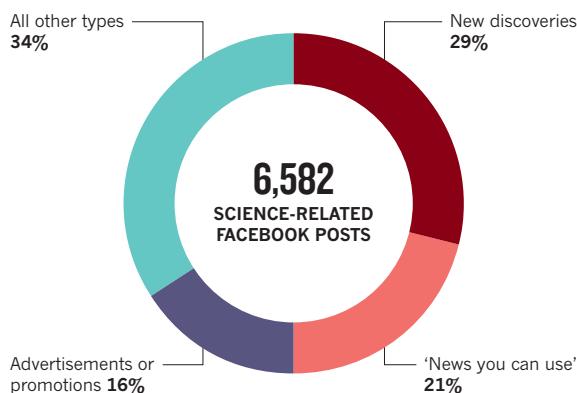
For daily news updates see:
www.nature.com/news

TREND WATCH

Facebook users routinely encounter science-related information in practical, 'news you can use' posts or in advertisements. A study by the Pew Research Center in Washington DC analysed 6,582 posts randomly selected from 30 science-related Facebook pages. Nearly 30% were directly about scientific discoveries, whereas 21% were 'news you can use' and 16% were promotions. The study found that visual posts, such as videos or images with little text, received most engagement.

SCIENCE ON SOCIAL MEDIA

An analysis of 30 science-related pages on Facebook found that almost one-third of the content covered new discoveries.



SOURCE: PEW RESEARCH CENTER

NEWS IN FOCUS

FORECASTING Researchers create tools to predict harmful plankton blooms **p.569**

AGEING Trial of diet restrictions in people reveals that they slow metabolism **p.570**

FUNDING US lawmakers approve biggest R&D budget in history **p.572**



DIVIDED OVER DNA Geneticists and archaeologists strive to find common ground **p.573**

GREG KENDALL-BALL/NATURE



A volunteer in the Alzheimer's study at the University of Antioquia, Colombia, who is unsure whether she carries a genetic mutation that causes the disease.

NEUROSCIENCE

Alzheimer's study zeroes in on enigmatic protein

Researchers in Colombia hope to uncover new drug targets.

BY SARA REARDON, MEDELLÍN, COLOMBIA

Jhon Kennedy was building a house for his family when he realized that his 45-year-old father was beginning to struggle with daily life. His father tried to help with the construction project but often forgot to complete simple tasks. And he kept getting lost on the way home from work.

Jhon Kennedy wasn't surprised: his four uncles had also started to lose their memories, one by one. But their doctors in Colombia's

rural Antioquia region, which is known for its mountainous terrain and coffee plantations, had never heard of early-onset dementia. It wasn't until a cousin learned about a study of Alzheimer's disease at the University of Antioquia in Medellín that Jhon Kennedy's relatives understood the illness they faced. For more than three decades, researchers there have been tracking a genetic mutation — common in the region — that causes Alzheimer's to strike people in their 40s and 50s.

Later this year, a team at the university will

begin scanning the brains of some of the study participants using a technique that is available only in a few major medical centres worldwide. It will allow the researchers to track a protein called tau, which accumulates rapidly in the brains of people with Alzheimer's as symptoms begin to emerge. Watching tau form in real time could reveal the part it plays in the disease, says Francisco Lopera, the neurologist who is leading the research.

Many scientists have long thought that the disease is triggered by another protein, ►



The University of Antioquia Brain Bank stores brains from people who had Alzheimer's disease.

► amyloid, that builds up in the brains of people with Alzheimer's. But several drugs that reduce amyloid levels have failed to relieve the symptoms of the disease in clinical trials, increasing interest in the role of tau.

If all goes well, Lopera's team will soon be the first in Colombia with the ability to scan people's brains for tau. In February, he and his colleagues published the results of a pilot study in which they brought 24 people from Colombia to Boston, Massachusetts, and used positron-emission tomography (PET) to search their brains for tau. The researchers showed for the first time that tau begins to accumulate in the brains of people with the Antioquia mutation six years before they begin to show signs of disease¹.

"This is a very definitive paper," says Bruce Miller, a behavioural neurologist at the University of California, San Francisco. "I think it's another piece of evidence that tau is very important and has a strong correlation with clinical symptoms."

The genetic mutation that affected Jhon Kennedy's father and uncles is well known in the field of Alzheimer's research. It probably arrived in South America with Spanish conquerors 375 years ago, and now affects 25 extended families in Antioquia with some 5,000 members. Researchers have published dozens of papers about this group, including some of the clearest evidence that amyloid

plaques can accumulate in the brain decades before symptoms of Alzheimer's appear².

But because questions have emerged about amyloid's role in Alzheimer's, scientists have taken a closer look at tau. The protein normally helps to stabilize the structures that enable neurons to communicate with one another. People with Alzheimer's make too much dysfunctional tau, causing these structures to collapse into tangles. The amount of tau seems to increase in step with a person's symptoms.

In the past several years, researchers have developed radioactive biomarkers that allow them to detect tau in living people using PET. Studies have shown that tau accumulation in the brain's language centres correlates with speech problems, for instance³. Several research groups are beginning clinical trials of drugs that scrub tau from the brain, although this work is at an early stage.

Jhon Kennedy and his 11 siblings each have a 50% chance of inheriting the Alzheimer's mutation from their father. Eight have enrolled in Lopera's study and could be among the first people to undergo tau imaging in Colombia; none know whether they carry the mutation.

The study, which began in 2013, was designed to test whether crenezumab, a drug that clears amyloid plaques from the brain, could lessen symptoms of Alzheimer's. In the past five years, Lopera's team has recruited 252 participants who were between 30 and 60 years

old. Those who carry the Alzheimer's mutation will begin accumulating amyloid in their brains during their early 30s, on average.

Each participant in the study will receive infusions of crenezumab or a placebo every other week, for five years. Lopera's team is also testing their cognitive abilities, scanning their brains for amyloid and searching for blood proteins and other biomarkers that could be early indicators of disease.

The scientists hope to begin mapping the tau in participants' brains in the coming months, after they receive final permission from regulators to produce the crucial radioactive marker GPT1. When that happens, the team will join a handful of researchers worldwide that are using the technology. "It's quite remarkable that a population that lives in such a remote area is getting access to some of the most advanced technology for understanding Alzheimer's," says Kenneth Kosik, a neuroscientist at the University of California, Santa Barbara.

Lopera and his colleagues want to determine how tau spreads in the brains of young people with Alzheimer's, and whether that pattern mirrors the distribution of tau seen in elderly people with the disease. They hope to compare their results with data from two clinical trials of anti-amyloid drugs in the United States that have begun scanning participants' brains for tau. Lopera, Kosik and other scientists are also beginning to identify families in Colombia with different genetic mutations that cause dementia or neurological disorders linked to tau, in the hope of imaging their brains too.

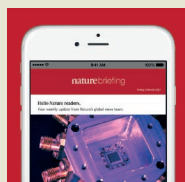
Lopera says that his team will not release data from the crenezumab study until it finishes in 2022. But if the results are promising, he adds, the researchers might give the drug to people under 30 who carry the mutation for early-onset Alzheimer's. The scientists have identified almost 500 young people who might be carriers, including Jhon Kennedy's 15-year-old daughter. "She's not worried about it," Jhon Kennedy says. "She's like me — she lives day by day." ■

1. Quiroz, Y. T. *et al.* *JAMA Neurol.* <http://dx.doi.org/10.1001/jamaneurol.2017.4907> (2018).
2. Reiman, E., *et al.* *Lancet Neurol.* **11**, 1048–56 (2012).
3. Ossenkoppele, R. *et al.* *Brain.* **139**, 1551–1567 (2016).

Travel for this story was supported by the Pulitzer Center on Crisis Reporting.

 **MORE ONLINE**

NATURE BRIEFING



Save time — get the daily *Nature* newsletter free in your inbox go.nature.com/2bhx2ks

MORE NEWS

- US science agencies set to win big in budget deal go.nature.com/2dv0wyc
- Early birds may have been too hefty to sit on their eggs go.nature.com/2gwama5
- HIV researcher is new head of US health agency go.nature.com/2pbabvn

NATURE PODCAST



Artificial intelligence in chemistry, and liquid droplets in living cells nature.com/nature/podcast

MATERIALS SCIENCE

Composite printer paves way for custom robots

Four-in-one device mixes materials with speed and precision.

BY MARK ZASTROW

A prototype 3D printer has for the first time combined several printing methods to enable researchers to produce devices from multiple materials in a single print run. So far, the machine has created basic electronic devices, but the technology brings materials scientists a step closer to their goal of printing complex equipment such as robots or smartphones.

The printer was presented at a meeting of the American Chemical Society in New Orleans, Louisiana, on 21 March.

"This is a remarkable technological advance and a great leap for the field of 3D printing," says Xuanhe Zhao, a materials scientist at the Massachusetts Institute of Technology in Cambridge, who was not involved in the work.

The most common 3D printers heat a plastic filament and lay it down in repeated lines, building a layered structure from the bottom up. This is the technique used in inexpensive consumer models. Several other 3D-printing methods have also emerged in recent years; these include spraying fine streams of aerosols, printing with liquid resin that is then cured to

form a flexible polymer, laying down thin layers of ink that are dried and hardened when exposed to light, and even printing with ink that contains conductive nanoparticles, to produce wires and circuits.

"Each printing technology has its own limitations," says Jerry Qi, a materials scientist at the Georgia Institute of Technology in Atlanta, who led the design of the multi-material printer. "We put four 3D-printing technologies under one platform."

Although current 3D printers can already produce electronic parts and devices made of multiple materials, if a structure requires more than one printing method a different machine is required for each. But moving an object from one printer to another is usually impractical for the micrometre-level precision that is required in 3D printing, and is inefficient if several materials are used in a single layer, says Qi.

His team's multimaterial printer has printheads — the nozzles that produce the

"This is a remarkable technological advance and a great leap for the field."

material — for each of the four techniques on a single printing platform. Each has its own software, lights for curing the materials, and a moving platform and robotic arms that can pick up and place components. This allows the printheads to work together to build single layers with a range of materials. "It is a very smart solution to this challenge," says Zhou.

Qi and his team have used the printer to embed a light-emitting diode inside a plastic case, printing the inner circuitry at the same time as the outer enclosure. They have also printed a layer of conductive ink inside a rubbery material that can stretch while delivering a current, demonstrating its potential for flexible electronics.

The printer has excited researchers in the field. "We just had a conversation where we were dreaming of that kind of machine," says Geoff Spinks, a materials engineer at the University of Wollongong in Australia. Such stretchable electronics could be used in compact soft robots, providing flexibility that is impossible using current methods of embedding wires into a device, he says.

The machine was funded mostly by the US Air Force and cost roughly US\$350,000 to build, says Qi. He thinks the printer's first customers will be aerospace companies that will use it to design circuitry for avionics, where printing antennas directly could enable rapid prototyping of avionic devices. Qi estimates the printer would sell for about \$1 million.

Spinks says this is only the beginning for multimaterial printers. He expects devices with expanded capabilities will soon become available: "I'd imagine in the near future we might have 16 different types of printheads." ■

ENVIRONMENT

Forecasting efforts target harmful plankton blooms

Ocean scientists seek automated system to predict the growth of tiny marine denizens.

BY JEFF TOLLEFSON

Massive blooms of a marine organism called *Noctiluca scintillans* are currently threatening fisheries, tourism and desalination plants in the Arabian Sea. But scientists are developing forecasting models that could help Oman and other countries in the region to predict when conditions seem ripe for a boom in these single-celled creatures, so that they can be better prepared.

The goal is to produce a mostly automated system that works like a weather forecast. It would ingest data on atmospheric and ocean conditions, and kick out regular seven-day bloom forecasts — which would be available to businesses, governments and scientists.

Such a system could also help researchers to understand the spread of plankton blooms in the Arabian Sea and beyond in the decades to come, says Joaquim Goes, a biological oceanographer at the Lamont-Doherty Earth Observatory in Palisades, New York,

and leader of the modelling project. The blooms might provide a short-term boost to some commercial fisheries, but marine biodiversity could ultimately decline if *Noctiluca* continues to grow and dominate the ecosystem, he says.

The project is one of several efforts around the world to develop models that can be used to understand and forecast harmful algal blooms. As early as this year, the US National Oceanic and Atmospheric Administration (NOAA) expects to launch its first

operational forecast for toxic plankton blooms in the Gulf of Maine using a similarly advanced model, says Richard Stumpf, a NOAA oceanographer in Silver Spring, Maryland.

Simulating, and then predicting, such blooms is difficult, Stumpf says, because researchers must understand not just local water chemistry and ocean circulation but also the behaviour of different plankton species.

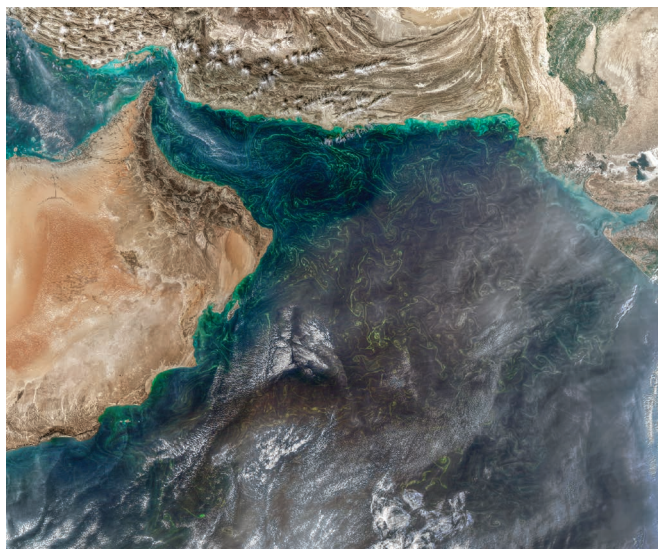
BOOM AND BUST

A huge bloom in the Arabian Sea that began last November is only now winding down. At its peak in January, it covered an area three times the size of Texas. “*Noctiluca* is completely overwhelming the system,” says Goes.

Goes and his team spent five weeks in Oman in January and February this year collecting samples and talking to local fishermen, as well as industry and government officials. They discussed the forecasting system and the kind of bloom-prediction information community members would need.

The team’s current forecasting model is promising, and the researchers hope to improve its resolution and to incorporate information on *Noctiluca*’s tricky biology.

This minute organism has both plant and animal characteristics. It is a predator that thrives in the low-oxygen waters that



Noctiluca blooms (green swirls) in the Arabian Sea, as seen from space.

have become more common in the region (H. R. Gomes *et al. Nature Commun.* 5, 4862; 2014). But *Noctiluca* also has a symbiotic relationship with algae that live inside the cells. These algae contain the green pigment chlorophyll, which sensors on satellites can monitor from space, enabling researchers to track the blooms.

A GROWING PROBLEM

It’s unclear why major *Noctiluca* blooms now occur in oceans around the world, including off the coasts of India, Thailand and Indonesia. Researchers suspect that the release

of raw sewage, which provides nutrients, and global warming might have roles.

The right conditions can stimulate *Noctiluca* blooms so thick that when they die, the decomposing bodies suck most of the oxygen out of the seawater. In years past, the bloom’s expanding low-oxygen zone drove sardines onto some beaches in Oman; aquaculture farms have also been affected.

This year, desalination plants had to scale back their activities to prevent *Noctiluca* from physically clogging their systems.

The current model devised by Goes and his team for the Indian Ocean has a resolution of 3.5 kilometres. It’s able to forecast the bloom potential of two types each of plant and animal plankton.

The model reproduces the large-scale behaviour of the current bloom in the Arabian Sea fairly well, but it’s not yet detailed enough to produce a useful forecast for a local shrimp farm or a desalination plant, says Sergio deRada, an ocean modeller at the Naval Research Laboratory at the Stennis Space Center in Mississippi.

The next step is to boost resolution to 1 kilometre or even 500 metres, and to represent *Noctiluca* itself in the model. The problem, says deRada, is that *Noctiluca*’s behaviour is more complex than that of the plankton currently included in the model. “That’s why modelling it is going to be tough.” ■

AQUA/MODIS/GSFC/NASA

MEDICINE

Reduced-calorie diet shows signs of slowing ageing

Restricting food intake dials down people’s metabolism in clinical trial.

BY ALISON ABBOTT

A study of people who reduced the calories they consumed has found the strongest evidence yet that such restrictions can slow down human metabolism. The results raise hopes that a low-calorie lifestyle — or treatments that mimic the biological effects of restricted eating — could prolong health in old age, and even extend life.

Past work in many short-lived animals, including worms, flies and mice, has shown that calorie restrictions reduce metabolism and extend lifespan. But experiments in

longer-living humans and other primates are more difficult to conduct and have not yet drawn clear conclusions.

The study was part of the multi-centre trial called CALERIE (Comprehensive Assessment of Long term Effects of Reducing Intake of Energy), sponsored by the US National Institutes of Health. The randomized, controlled trial tested the effects of 2 years of caloric restriction on metabolism in more than 200 healthy, non-obese adults.

“The CALERIE trial has been important in addressing the question of whether the pace of ageing can be altered in humans,”

says Rozalyn Anderson, who studies ageing at the University of Wisconsin–Madison. She leads one of two large, independent studies on calorie restriction in rhesus monkeys. “This new report provides the most robust evidence to date that everything we have learnt in other animals can be applied to ourselves.”

Published on 22 March in *Cell Metabolism*, the latest study looked at 53 CALERIE participants who had been recruited at the Pennington Biomedical Research Center in Baton Rouge, Louisiana (L. M. Redman *et al. Cell Metab.* <http://doi.org/cmr; 2018>). This

EXTRASOLAR PLANETS

Green light for exoplanet probe

Europe's ARIEL mission will study planetary atmospheres.

BY ELIZABETH GIBNEY

The world's first space telescope dedicated to studying the atmospheres of exoplanets is set to launch in 2028, after the European Space Agency (ESA) selected the mission for development on 20 March.

The Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL) — to be built with a budget of €450 million (US\$558 million) — will chart more than 1,000 known planetary systems outside our own. Its mission is to understand the links between a planet's chemistry and its environment, says Giovanna Tinetti, a planetary scientist at University College London and principal investigator for the mission. So far, exoplanet probes have mainly focused on finding new planets and measuring their sizes, masses and orbits. "ARIEL can really give us a full picture of what exoplanets are made of, how they form and how they evolve," says Tinetti.

A spectrograph aboard the observatory will study the light that filters through a planet's atmosphere as it passes across the face of its host star, revealing chemical fingerprints of gases that shroud the body. The mission will also try to refine estimates of a planet's temperature by teasing out how light from its star changes when the body moves behind it, revealing details about a planet's overall radiation. Tinetti says that it will monitor the worlds at snapshots in time, and also measure how the chemistry and atmospheres of planets vary across their surfaces and with the seasons.

Since the first exoplanet discoveries in the 1990s, some 3,700 worlds have been confirmed in 2,800 systems. ARIEL will focus on planets that have an estimated temperature of more than 350°C — and are thus unlikely to host life as we know it — and a range of masses, in particular those heavier than a few Earth masses. The observations will build a 'standard model' of how a planet's chemistry depends on its star and the conditions of its birth, Tinetti says.

ARIEL is one of several exoplanet missions in the works around the world, but its focus on planetary atmospheres is rare. Other missions are studying planet size, density and mass. Together, this generation of telescopes will allow scientists to test models of planetary formation and composition, says Elizabeth Tasker, an exoplanet researcher at the Japan Aerospace Exploration Agency's Institute of Space and Astronautical Science in Sagami-hara. ■

CARLOS SANCHEZ



Reduced food intake has affected people's metabolism in a controlled study.

facility is home to 4 of the world's 20 or so state-of-the-art metabolic chambers, which are like small, sealed hotel rooms that measure minute-by-minute the amount of oxygen that occupants use and how much carbon dioxide they exhale. This allows researchers to track how the occupants use energy with unprecedented precision, says Anderson. The ratio between the two gases, combined with analysis of nitrogen in occupants' urine, indicates whether the occupant is burning fat, carbohydrate or protein.

The trial participants, aged between 21 and 50, were randomized into two groups: 34 reduced their calorie intake by an average of 15%, and 19 ate as usual. At the end of each of the two years, they all underwent tests related to metabolism and biological markers of ageing. They were also placed in a metabolic chamber for 24 hours.

The scientists found that participants on the diet used energy much more efficiently while sleeping than did the control group. This reduction in their base metabolic rate was greater than would be expected as a result of the test group's weight loss, which averaged nearly 9 kilograms per participant. All the other clinical measurements were in line with reduced metabolic rate, and indicated a decrease in damage due to ageing.

MODEL METABOLISM

Caloric restriction has been known for decades to extend life in different species. In the 1990s, scientists began to identify the genes and biochemical pathways that are actively involved in longevity in the short-lived worm *Caenorhabditis elegans*, and in the fly *Drosophila melanogaster*. These include pathways relevant to insulin sensitivity and the function of mitochondria — cellular structures that use oxygen to generate energy. Later studies revealed that calorie restrictions alter similar pathways in mice

and monkeys. Mice on restricted diets can live up to 65% longer than mice allowed to eat freely, and the ongoing monkey studies hint at longer survival and reduced signs of ageing.

"The Rolls-Royce of a human longevity study would carry on for many decades to see if people do actually live longer," says Pennington physiologist Leanne Redman, the lead author of the latest study. CALERIE ran for two years, and aimed to see whether a calorie-restricted diet in humans induces some of the same metabolic, hormonal and gene-expression adaptations that are thought to be involved in slowing ageing in other species during long-term caloric restriction.

Few people would want, or be able, to restrict their diet as severely as the participants in the study. "But understanding the biology of how restricting calories extends life will allow us to find easier ways to intervene," says Anderson.

Redman would like to repeat the study, combining less-ambitious calorie restriction with a diet containing antioxidant food to reduce damage from oxygen free radicals, or with a drug such as resveratrol, which mimics key aspects of calorie restriction.

Other scientists are starting to test restricting calories for a few days every month. Such intermittent limitations have been found to be as effective as continuous calorie restriction in protecting mice against diseases of ageing such as diabetes and neurodegeneration (V. D. Longo and M. P. Mattson *Cell Metab.* **19**, 181–192; 2014). "I think that's going to be a way to get all the benefits, without the problems of constant dieting," says gerontologist Valter Longo of the University of Southern California in Los Angeles, who is embarking on clinical trials of intermittent calorie restriction in various disorders. ■

POLITICS

US science agencies reap budget windfall

Government spending on research and development rises to a historic high in 2018.

BY LAUREN MORELLO AND
GEORGIA GUGLIELMI

Nearly all US science agencies will see their budgets grow in 2018, under a US\$1.3-trillion spending deal enacted on 23 March. For the second year in a row, lawmakers in Congress ignored the steep cuts sought by President Donald Trump.

The legislation boosts funding for the National Institutes of Health (NIH) to \$37 billion, \$3 billion over the 2017 level. The National Science Foundation receives \$7.8 billion, \$295 million more than it got last year. And NASA's budget rises to \$20.7 billion, an increase of \$1.1 billion.

One notable outlier to the overall trend is the Environmental Protection Agency, whose funding will stay frozen at the 2017 level of \$8.1 billion. But that might be a victory of sorts, given that Trump had proposed slashing the agency's budget by more than 30% in 2018, to \$5.7 billion.

"This is a good deal," says Benjamin Corb, director of public affairs at the American Society for Biochemistry and Molecular Biology in Rockville, Maryland. "We're happy to see these numbers" (see 'Budget bonanza'). The budget legislation lifts overall US government spending on research and development to a historic high in 2018, according to an analysis by the American Association for the Advancement of Science in Washington DC.

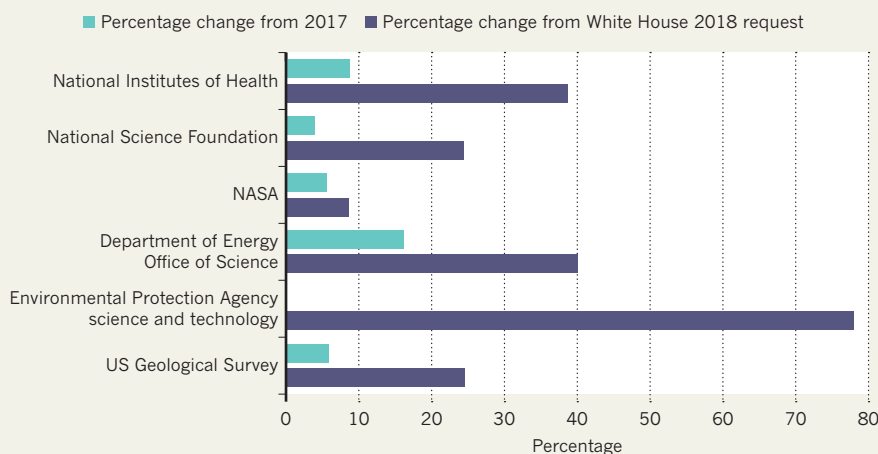
The deal includes increases for several of the NIH's signature programmes. The BRAIN initiative (Brain Research through Application of Innovative Neurotechnologies) receives \$400 million, up \$140 million from 2017. And the All of Us research programme — a 10-year effort to track the health of 1 million Americans, which was formerly known as the Precision Medicine Initiative — gets \$290 million. That is a \$60-million boost over last year.

GUN DEBATE

The Centers for Disease Control and Prevention (CDC) receives \$8.3 billion, an increase of \$1.1 billion over the 2017 level. Lawmakers in Congress reportedly debated whether to abandon a long-standing provision that prohibits the CDC from advocating or promoting gun control, which has been interpreted widely as a ban on gun research funded or conducted

BUDGET BONANZA

The US Congress negotiated a 2018 spending deal that ignores the steep cuts to science funding sought by President Donald Trump.



by the CDC. But the spending bill stops short of repealing the provision, known as the Dickey Amendment.

Instead, an official report accompanying the legislation notes that, although it includes language to prohibit the CDC and other agencies from funding activities to advocate or promote gun control, "the Secretary of Health and Human Services has stated the CDC has the authority to conduct research on the causes of gun violence".

That "does nothing", says David Hemenway, an economist who directs the Harvard Injury

"This is a good deal. We're happy to see these numbers."

Control Research Center at Harvard University in Boston, Massachusetts. "Everybody knows that the Dickey Amendment does not prohibit federal funding of gun research." But officials at the CDC also know that if they fund gun research, "they're going to get beaten up".

The budget deal sets aside \$6.2 billion for NASA's science directorate, an increase of \$457 million from the 2017 level. That includes \$595 million to send an orbiter and lander to Jupiter's moon Europa; Trump had sought \$425 million to send a spacecraft flying past that moon. The legislation also includes funding for four Earth-science programmes that Trump

had sought to cut, including the Orbiting Carbon Observatory-3, which would measure atmospheric carbon dioxide from space.

Lawmakers also rejected Trump's plan to cancel the Wide-Field Infrared Survey Telescope (WFIRST), which would hunt exoplanets and dark matter — a project that US astronomers ranked as their top priority in a 2010 survey of science priorities for the next decade.

The funding deal includes \$6.2 billion for the Department of Energy's Office of Science, an \$868-million jump from the 2017 level. Trump had sought to cut its budget to just under \$4.5 billion. Congress also rejected the president's proposal to slash funding for the Advanced Research Projects Agency — Energy (ARPA-E), which pursues risky research that could yield major advances. The spending legislation includes \$353 million for ARPA-E; Trump had proposed \$20 million.

Despite the windfall for science agencies, some budget experts warn that science advocates should not become complacent. "Don't look for a repeat of this next year," says Michael Lubell, a physicist at the City College of New York. The large increases for science agencies came after Congress moved in February to relax mandatory spending caps for 2018 and 2019. That added \$63 billion to the overall funding pot this year, but the number will grow only slightly in 2019, to \$68 billion. ■

SOURCE: AAAS



THE BATTLE FOR COMMON GROUND

Ancient genomes are revolutionizing the study of human prehistory but sometimes straining the relationships between archaeologists and geneticists.

BY EWEN CALLAWAY

Thirty kilometres north of Stonehenge, through the rolling countryside of south-west England, stands a less-famous window into Neolithic Britain. Established around 3600 BC by early farming communities, the West Kennet long barrow is an earthen mound with five chambers, adorned with giant stone slabs. At first, it served as a tomb for some three dozen men, women and children. But people continued to visit for more than 1,000 years, filling the chambers with relics such as pottery and beads that have been interpreted as tributes to ancestors or gods.

The artefacts offer a view of those visitors and their relationship with the wider world. Changes in pottery styles there sometimes

echoed distant trends in continental Europe, such as the appearance of bell-shaped beakers — a connection that signals the arrival of new ideas and people in Britain. But many archaeologists think these material shifts meshed into a generally stable culture that continued to follow its traditions for centuries.

“The ways in which people are doing things are the same. They’re just using different material culture — different pots,” says Neil Carlin at University College Dublin, who studies Ireland and Britain’s transition from the Neolithic into the Copper and Bronze Ages.

But last year, reports started circulating that seemed to challenge this picture of stability. A study¹ analysing genome-wide

The West Kennet long barrow served as a tomb and ceremonial site for more than a millennium.

data from 170 ancient Europeans, including 100 associated with Bell Beaker-style artefacts, suggested that the people who had built the barrow and buried their dead there had all but vanished by 2000 BC. The genetic ancestry of Neolithic Britons, according to the study, was almost entirely displaced. Yet somehow the new arrivals carried on with many of the Britons' traditions. "That didn't fit for me," says Carlin, who has been struggling to reconcile his research with the DNA findings.

The Bell Beaker 'bombshell' study appeared in *Nature*² in February and included 230 more samples, to make it the largest ancient-genome study on record. But it is just the latest example of the disruptive influence that genetics has had on the study of the human past. Since 2010, when the first ancient-human genome was fully sequenced³, researchers have amassed data on more than 1,300 individuals (see 'Ancient genomes'), and used them to chart the emergence of agriculture, the spread of languages and the disappearance of pottery styles — topics that archaeologists have laboured over for decades.

Some archaeologists are ecstatic over the possibilities offered by the new technology. Ancient-DNA work has breathed new life and excitement into their work, and they are beginning once-inconceivable investigations, such as sequencing the genome of every individual from a single graveyard. But others are cautious.

"Half the archaeologists think ancient DNA can solve everything. The other half think ancient DNA is the devil's work," quips Philipp Stockhammer, a researcher at Ludwig-Maximilians University in Munich, Germany, who works closely with geneticists and molecular biologists at an institute in Germany that was set up a few years ago to build bridges between the disciplines. The technology is no silver bullet, he says, but archaeologists ignore it at their peril.

Some archaeologists, however, worry that the molecular approach has robbed the field of nuance. They are concerned by sweeping DNA studies that they say make unwarranted, and even dangerous, assumptions about links between biology and culture. "They give the impression that they've sorted it out," says Marc Vander Linden, an archaeologist at the University of Cambridge, UK. "That's a little bit irritating."

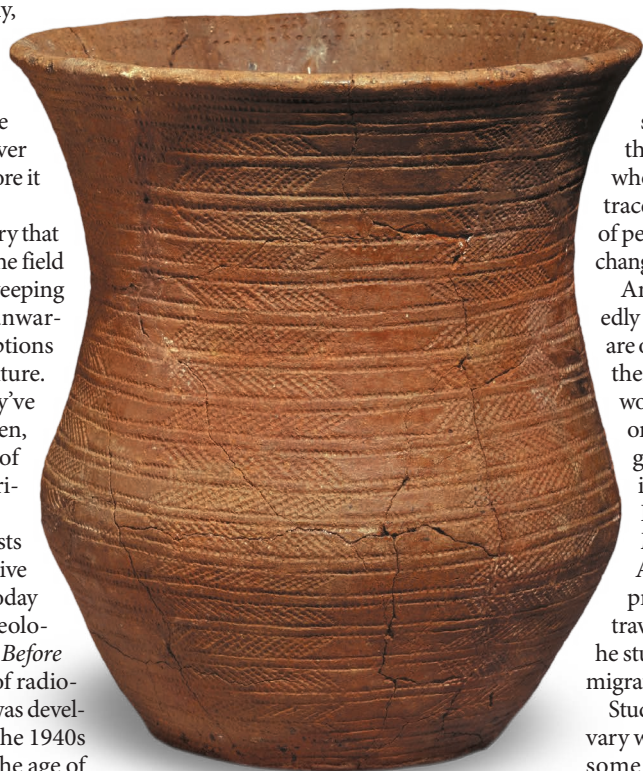
This isn't the first time archaeologists have had to contend with transformative technology. "The study of prehistory today is in crisis," wrote Cambridge archaeologist Colin Renfrew in his 1973 book *Before Civilization*, describing the impact of radiocarbon dating. Before the technique was developed by chemists and physicists in the 1940s and 50s, prehistorians determined the age of sites using 'relative chronologies', in some cases relying on ancient Egyptian calendars and false

assumptions about the spread of ideas from the Near East. "Much of prehistory, as written in the existing textbooks is inadequate: some of it, quite simply wrong," Renfrew surmised.

It wasn't an easy changeover — early carbon-dating efforts were off by hundreds of years or more — but the technique eventually allowed archaeologists to stop spending most of their time worrying about the age of bones and artefacts and focus instead on what the remains meant, argues Kristian Kristiansen, who studies the Bronze Age at the University of Gothenburg in Sweden. "Suddenly there was a lot of free intellectual time to start thinking about prehistoric societies and how they are organized." Ancient DNA now offers the same opportunity, says Kristiansen, who has become one of his field's biggest cheerleaders for the technology.

Genetics and archaeology have been uneasy

"THESE RESULTS WERE A SHOCK TO THE ARCHAEOLOGICAL COMMUNITY."



Bell Beaker pots signal a period of unprecedented cultural intermingling for early Europeans.

bedfellows for more than 30 years — the first ancient-human DNA paper⁴, in 1985, reported sequences from an Egyptian mummy (now thought to be contamination). But improvements in sequencing technology in the mid-to-late 2000s set the fields on a collision course.

In 2010, scientists led by Eske Willerslev at the Natural History Museum of Denmark used DNA from a lock of hair from a 4,000-year-old native Greenlander to generate the first complete sequence of an ancient-human genome³. Seeing the future of the field before his eyes, Kristiansen asked Willerslev to team up on a prestigious European Research Council grant that would allow them to examine human mobility as the late Neolithic gave way to the Bronze Age, some 4,000–5,000 years ago.

ASSOCIATION PROBLEMS

Migration has been a major source of tension for archaeologists. They have debated at length whether human movements are responsible for cultural changes in the archaeological record, such as the Bell Beaker phenomenon, or whether it is simply the ideas that are moving through cultural exchanges. Populations identified by the artefacts they associated with came to be seen as a remnant of the science's colonial past, and one that imposed artificial categories. "Pots are pots, not people," goes a common refrain.

Most archaeologists have since cast aside the view that prehistory was like a game of *Risk*, in which homogenous cultural groups conquer their way across a map of the world. Instead, researchers tend to focus on under-

standing a small number of ancient sites and the lives of the people who lived there. "Archaeology had moved away from these grand narratives," says Tom Booth, a bioarchaeologist at the Natural History Museum in London, who is part of a team using ancient DNA to trace the arrival of farming in Britain. "A lot of people thought you needed to understand change regionally to understand people's lives."

Ancient-DNA work — which has repeatedly shown that a region's modern inhabitants are often distinct from populations that lived there in the past — promised, for better or worse, to bring back some of the broad focus on migration to human prehistory. "What genetics is particularly good at is detecting change in populations," says David Reich, a population geneticist at Harvard Medical School in Boston, Massachusetts. Archaeologists, Kristiansen says, "were prepared to accept that individuals had travelled". But for the Bronze Age period that he studies, "they were not prepared for major migrations. That was a new thing."

Studies of strontium isotopes in teeth⁵, which vary with local geochemistry, had hinted that some Bronze Age individuals had moved hundreds of kilometres over their lifetimes, Kristiansen says. He and Willerslev wondered

SOURCE: DAVID REICH

whether DNA analysis might detect movements of whole populations during this period.

They would have competition. In 2012, David Anthony, an archaeologist at Hartwick College in Oneonta, New York, loaded his car with boxes of human remains that he and his colleagues had excavated from the steppes near the Russian city of Samara, including bones associated with a Bronze Age pastoralist culture called the Yamnaya. He was bringing them to the ancient-DNA lab just established by Reich in Boston. Like Kristiansen, Anthony was comfortable theorizing about the past on a grand scale. His 2007 book *The Horse, the Wheel and Language* proposed that the Eurasian steppe had been a melting pot for the modern developments of horse domestication and wheeled transport, which propelled the spread of a family of languages called Indo-European across Europe and parts of Asia.

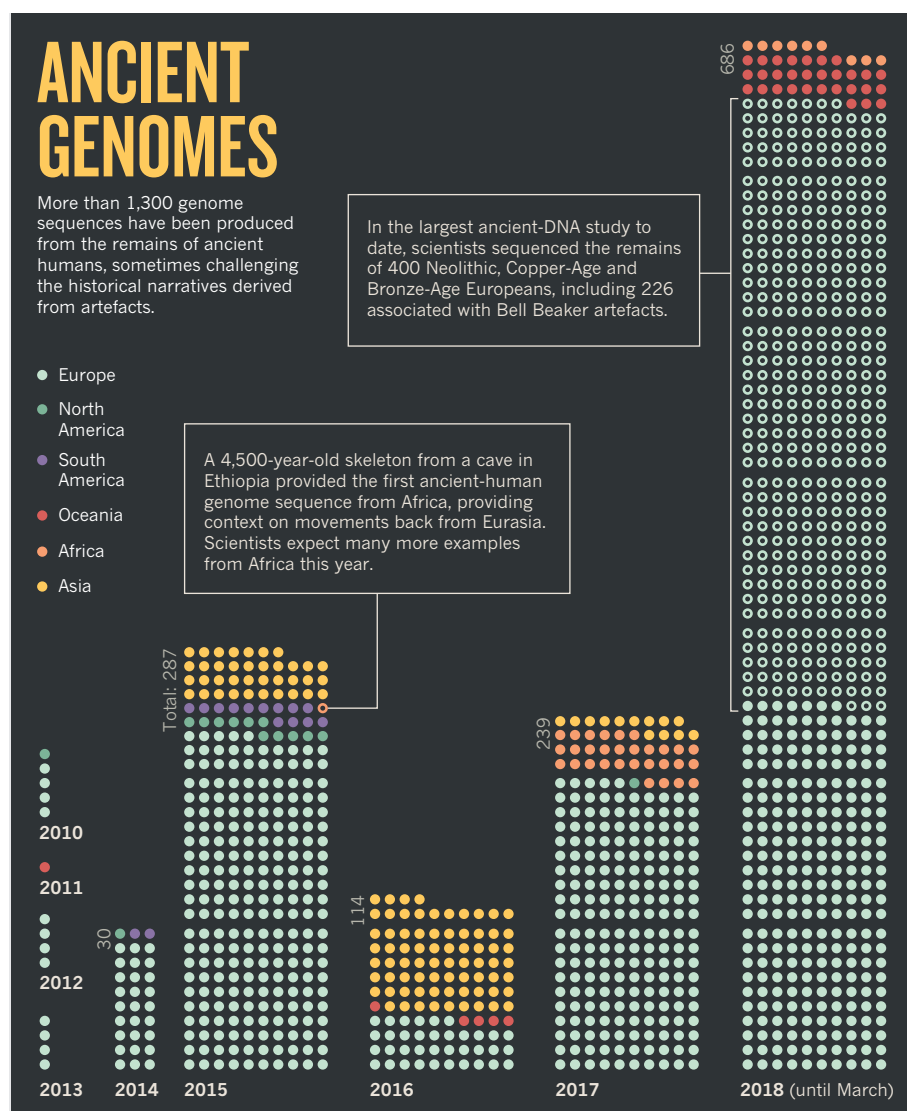
In duelling 2015 *Nature* papers^{6,7}, the teams arrived at broadly similar conclusions: an influx of herders from the grassland steppes of present-day Russia and Ukraine — linked to Yamnaya cultural artefacts and practices such as pit burial mounds — had replaced much of the gene pool of central and Western Europe around 4,500–5,000 years ago. This was coincident with the disappearance of Neolithic pottery, burial styles and other cultural expressions and the emergence of Corded Ware cultural artefacts, which are distributed throughout northern and central Europe. “These results were a shock to the archaeological community,” Kristiansen says.

CORD CUTTERS

The conclusions immediately met with push-back. Some of it began even before the papers were published, says Reich. When he circulated a draft among his dozens of collaborators, several archaeologists quit the project. To many, the idea that people linked to Corded Ware had replaced Neolithic groups in Western Europe was eerily reminiscent of the ideas of Gustaf Kossinna, the early-twentieth-century German archaeologist who had connected Corded Ware culture to the people of modern Germany and promoted a ‘Risk board’ view of prehistory known as settlement archaeology. The idea later fed into Nazi ideology.

Reich won his co-authors back by explicitly rejecting Kossinna’s ideas in an essay included in the paper’s 141-page supplementary material⁷. He says the episode was eye-opening in showing how a wider audience would perceive genetic studies claiming large-scale ancient migrations.

Still, not everyone was satisfied. In an essay⁸ titled ‘Kossinna’s Smile’, archaeologist Volker Heyd at the University of Bristol, UK, disagreed, not with the conclusion that people moved west from the steppe, but with how their genetic signatures were conflated with complex cultural expressions. Corded Ware and Yamnaya burials are more different than they are similar, and there is evidence of cultural exchange, at least, between the Russian steppe and regions



west that predate Yamnaya culture, he says. None of these facts negates the conclusions of the genetics papers, but they underscore the insufficiency of the articles in addressing the questions that archaeologists are interested in, he argued. “While I have no doubt they are basically right, it is the complexity of the past that is not reflected,” Heyd wrote, before issuing a call to arms. “Instead of letting geneticists determine the agenda and set the message, we should teach them about complexity in past human actions.”

Ann Horsburgh, a molecular anthropologist and prehistorian at Southern Methodist University in Dallas, Texas, attributes such tensions to communication problems. Archaeology and genetics say distinct things about the past, but often use similar terms, such as the name of a material culture. “It’s C. P. Snow all over again,” she says, referring to the influential ‘Two Cultures’ lectures by the British scientist lamenting the deep intellectual divide between the sciences and the humanities. Horsburgh complains that genetic results are too often given precedence over inferences about the past from archaeology and anthropology, and

that such “molecular chauvinism” prevents meaningful engagement⁹. “It’s as though genetic data, because they’re generated by people in lab coats, have some sort of unalloyed truth about the Universe.”

Horsburgh, who is seeing her own field of African prehistory start to feel the tremors of ancient genomics, says that archaeologists frustrated at having their work misinterpreted should wield their power over archaeological remains to demand more equitable partnerships with geneticists. “Collaboration doesn’t mean I send you an e-mail saying ‘hey, you’ve got some really cool bones. I’ll get you a *Nature* paper.’ That’s not a collaboration,” she says.

Many archaeologists are also trying to understand and engage with the inconvenient findings from genetics. Carlin, for instance, says that the Bell Beaker genome study sent him on “a journey of reflection” in which he questioned his own views about the past. He has pored over the selection of DNA samples included in the study as well as the basis for its conclusion that the appearance of Bell Beaker artefacts coincided with a greater than 90% replacement in

Britain's gene pool. "I didn't want to be questioning it from a position of ignorance," Carlin says.

Like Heyd, he accepts that a shift in ancestry occurred (although he has questions about its timing and scale). Those results, in fact, now have him wondering about how cultural practices such as leaving pottery and other tributes at the West Kennet long barrow persisted in the face of such upheavals. "I would characterize a lot of these papers as 'map and describe.' They're looking at the movement of genetic signatures, but in terms of how or why that's happening, those things aren't being explored," says Carlin, who is no longer disturbed by the disconnect. "I am increasingly reconciling myself to the view that archaeology and ancient DNA are telling different stories." The changes in cultural and social practices that he studies might coincide with the population shifts that Reich and his team are uncovering, but they don't necessarily have to. And such biological insights will never fully explain the human experiences captured in the archaeological record.

Reich agrees that his field is in a "map-making phase", and that genetics is only sketching out the rough contours of the past. Sweeping conclusions, such as those put forth in the 2015 steppe migration papers, will give way to regionally focused studies with more subtlety.

This is already starting to happen. Although the Bell Beaker study found a profound shift in the genetic make-up of Britain, it rejected the notion that the cultural phenomenon was associated with a single population. In Iberia, individuals buried with Bell Beaker goods were closely related to earlier local populations and shared little ancestry with Beaker-associated individuals from northern Europe (who were related to steppe groups such as the Yamnaya). The pots did the moving, not the people.

Reich describes his role as that of a 'mid-wife' delivering ancient-DNA technology to archaeologists, who can apply it as they see fit. "Archaeologists will embrace this technology and will not be Luddites," he predicts, "and they'll make it their own."

A STRONGER PARTNERSHIP

Nestled in a sleepy valley in the state of Thuringia in former East Germany, the city of Jena has become an unlikely hub for the convergence of archaeology and genetics. In 2014, the prestigious Max Planck Society established an Institute for the Science of Human History there and installed a rising star in ancient-DNA research, Johannes Krause, as a director. Krause was a protégé of the geneticist Svante Pääbo, at the Max Planck Institute for Evolutionary Anthropology in Leipzig. There, Krause worked on the Neanderthal genome¹⁰ and helped discover a new archaic human group, known as Denisovans¹¹.

Whereas Pääbo was focused on applying genetics to biological questions about ancient humans and their relatives, Krause saw a wider scope for the technology. Before leading the Jena

institute, his team identified DNA from plague-causing bacteria in the teeth of people who died from the Black Death in the fourteenth century, the first direct evidence of a potential cause for the pandemic¹². At Jena, Krause hoped to bring genetics to bear, not just on 'prehistorical' periods such as the Neolithic and the Bronze Age, where archaeological methods are the main tool for reconstructing the past, but also on more-recent times. Outreach with historians is still a work in progress, but archaeology and genetics are thoroughly embedded at the institute. The department Krause directs is even called archaeogenetics. "We have to be interdisciplinary," he says, because geneticists are addressing

"ARCHAEOLOGISTS WILL EMBRACE THIS TECHNOLOGY AND WILL NOT BE LUDDITES."

questions and time periods that archaeologists, linguists and historians have been poring over for decades.

Krause and his team have been heavily involved in the map-making phase of ancient genomics (he worked closely with Reich's team on many such projects). But a study published late last year¹³ that focused on the transition between the Neolithic and Bronze Age in Germany won plaudits from archaeologists who have been dubious of the larger-scale ancient-DNA studies.

Led by Stockhammer, who also has a post at the Jena institute, the team analysed 84 Neolithic and Bronze Age skeletons from southern Bavaria's Lech River Valley dating to between 2500 and 1700 BC. The diversity in the genomes of cellular structures known as mitochondria, which are inherited maternally, rose during this period, suggesting an influx of women. Meanwhile, strontium isotope levels in teeth — which are set during childhood — suggested that most females weren't local. In one case, two related individuals who lived within a few generations of each other were found buried with different material cultures. In other words, some cultural shifts in the archaeological record could be due not to massive migrations, but to the systematic mobility of individual women.

It is the prospect of more such studies that has archaeologists salivating over ancient DNA. In the near future, says Stockhammer, archaeologists will be able to sequence the genomes of all the individuals at a burial site and build a

local family tree, while also determining how individuals fit into larger ancestry patterns. This should allow researchers to ask how biological kinship relates to the inheritance of material culture or status. "These are the big questions of history. They can be solved now only with collaboration," says Stockhammer.

Another glimpse of this approach appeared in February on the bioRxiv preprint server¹⁴. The paper explores Europe's migration period, when 'barbarian hordes' filled the void left after the fall of the Roman Empire. In the paper, a team of geneticists, archaeologists and historians built family trees of 63 individuals from two medieval cemeteries in Hungary and northern Italy associated with a group known as the Longobards. They found evidence of high-status outsiders buried in the cemetery: most bore central and northern European genetic ancestry that differed from that of local people, who tended to be buried without goods — offering tentative support to the idea that some barbarian groups included outsiders.

Patrick Geary, a medieval historian at the Institute for Advanced Study in Princeton, New Jersey, who co-led the Longobard study, would not comment on the research because it is now being peer reviewed. But he says that genetic studies of historical times, such as the migration period, carry pitfalls, too. Historians are increasingly incorporating data such as palaeoclimate records into their work, and will do likewise with ancient DNA, Geary says. But they share archaeologists' fears that biology and culture will be conflated, and that problematic designations such as Franks or Goths or Vikings will be reified by genetic profiles, overriding insights into how ancient peoples viewed themselves. "These days, what historians want to know about is identity," he says. "Genetics cannot answer these questions."

Reich concedes that his field hasn't always handled the past with the nuance or accuracy that archaeologists and historians would like. But he hopes they will eventually be swayed by the insights his field can bring. "We're barbarians coming late to the study of the human past," Reich says. "But it's dangerous to ignore barbarians." ■ [SEE EDITORIAL P.559](#)

Ewen Callaway writes for *Nature* from London.

1. Olalde, I. *et al.* Preprint at bioRxiv <http://dx.doi.org/10.1101/135962> (2017).
2. Olalde, I. *et al.* *Nature* **555**, 190–196 (2018).
3. Rasmussen, M. *et al.* *Nature* **463**, 757–762 (2010).
4. Pääbo, S. *Nature* **314**, 644–645 (1985).
5. Frei, K. M. *et al.* *Sci. Rep.* **5**, 10431 (2015).
6. Allentoft, M. E. *et al.* *Nature* **522**, 167–172 (2015).
7. Haak, W. *et al.* *Nature* **522**, 207–211 (2015).
8. Heyd, V. *Antiquity* **91**, 348–359 (2017).
9. Horsburgh, K. A. J. *Archaeol. Sci.* **56**, 141–145 (2015).
10. Green, R. E. *et al.* *Science* **328**, 710–722 (2010).
11. Krause, J. *et al.* *Nature* **464**, 894–897 (2010).
12. Bos, K. I. *et al.* *Nature* **478**, 506–510 (2011).
13. Knipper, C. *et al.* *Proc. Natl Acad. Sci. USA* **114**, 10083–10088 (2017).
14. Amorim, C. E. G. *et al.* Preprint at bioRxiv <http://dx.doi.org/10.1101/268250> (2018).

COMMENT

LAB LIFE Freeman Dyson's letters reveal a delighted noticer of details **p.581**



PHYSICS A history of how quantum physics picked its winners **p.582**

SUSTAINABILITY Parsing the relationship between climate and conflict **p.587**

OBITUARY John Sulston, pioneer of the Human Genome Project and open data **p.588**

DAVID HOWELLS/CORBIS/GETTY



Cyberspy Shannen Rossmiller posed as male Iraqi and Afghan militants in extremist chat rooms to expose weapon caches and bomb plots.

Cybersecurity needs women

Safeguarding our lives online requires skills and experiences that lie beyond masculine stereotypes of the hacker and soldier, says **Winifred R. Poster**.

Computer hacking is becoming more widespread and damaging. Headlines highlight attacks on government agencies, political campaign offices, financial institutions and big corporations. But citizens and consumers are paying a heavy price. In 2016, 2 billion people had their personal details stolen, including the medical records of more than 100 million Americans. Hacks of US retail outlets such as Target and global credit companies such as Equifax compromised the private data of hundreds of millions of customers. In the past 6 years, more than US\$107 billion was stolen from US consumers through identity theft.

Cybercrime exacerbates inequalities. A

million more US women than men had their identities stolen in 2014. People of African American and Latino descent are, on average, two to three times more likely than white people to be victims of fraud related to debt or income. And women and girls are more likely than men to be targets of 'remote sexual abuse' — coerced into posing nude online or being stalked through the Internet.

Security technologies also disadvantage women and other groups. For example, biometric facial recognition systems have trouble identifying the faces of women and people of colour. Airport security systems and operators disproportionately flag black women for

strip searches relative to other passengers¹.

Cybersecurity professionals — who protect databases, software systems and computer networks from access, change or destruction — are predominantly male. Women comprise only 11% of these professionals worldwide, and only 14% in North America (see 'Women in cybersecurity'). By comparison, women make up 57% of the US workforce. Even cybersecurity's sister industries do better: 15% of the US military and 25% of staff in information technology are women (see 'Sister fields'). By 2020, 2 million more cybersecurity jobs will be needed worldwide in addition to the 3.2 million people who are already employed in the ►

► field, of whom almost 750,000 are in the United States.

Cybersecurity's future depends on its ability to attract, retain and promote women, who represent a highly skilled and under-tapped resource. The discipline also needs to learn about women's experiences as victims of cybercrime and the steps needed to address the imbalance of harm.

Here I highlight four ways in which the field should adapt.

FOUR PRIORITIES

Acknowledge women's contributions. Women have been working in cybersecurity for a century. Yet many of their stories have been sidelined because of the secrecy of the work, its wartime contexts or because male colleagues have been put in the limelight instead.

During the Second World War, the United States employed 10,000 women as 'code girls' to decipher encrypted messages sent by the Japanese and Germans. Likewise, the United Kingdom hired more than 7,000 women to work at Bletchley Park, its centre for cryptanalysis, where they made up about three-quarters of the workforce. And in 1942, actor and inventor Hedy Lamarr patented an encryption method for signalling to torpedoes that is now the basis of WiFi and bluetooth technology.

Elizbeth Smith Friedman helped to invent the science of cryptography for the US Federal Bureau of Investigation (FBI)². Her techniques in the 1940s broke international spy rings, decoded three Nazi Enigma machines and contributed to the early work of the forerunner to the Central Intelligence Agency. Yet after the war, her elite code-breaking unit was shut down and various men took credit for her work, including her husband William Friedman and FBI

director J. Edgar Hoover. Indeed, Hoover showed outright hostility towards women in this field — when he started as director of the FBI's predecessor in 1924, he fired all the female agents and banned further recruitment of women in these roles.

Women were the first programmers, calculating weapons trajectories by hand and entering them into the Electronic Numerical Integrator And Computer (ENIAC) at the University of Pennsylvania, Philadelphia, in the 1940s³. In fact, 'computer' originally referred not to the machine but to the women who programmed it. Other women developed the first programming languages, methods to detect intrusions into computer systems and network bridges between communication centres⁴. In the 1950s, African American female mathematicians at NASA calculated the aeronautical trajectories that put men on the Moon⁵. The proportion of women in computer science grew until the mid-1980s — the dawn of personal computing — when numbers dropped precipitously. Today, only around 18% of US computer-science majors are women, compared with around 37% in 1984.

Even so, in the past decade, women have held influential positions in US national cybersecurity. Theresa Payton was the first female chief information officer in the White House, under former president George W. Bush. Former president Barack Obama appointed Melissa Hathaway in 2009 as his first 'cyber czar', in the role of acting senior director for cyberspace for the National Security Council. Letitia Long was the first woman to head a major US intelligence agency, as director of the National

"Women have held influential positions in US national cybersecurity."

Geospatial-Intelligence Agency from 2010 to 2014. This supplied the satellite, geographical and social-media data that enabled Osama bin Laden's capture. Janet Napolitano headed the Department of Homeland Security (DHS) from 2009 to 2013, and Kirstjen Nielsen has been at the helm since 2017. Jeannette Manfra, the chief cybersecurity official for the DHS, is leading the investigation into Russian hacking of US voter registration rolls before the 2016 presidential election.

Research leadership is crucial because it dictates the direction of security technologies and strategies. From 2009 to 2012, Regina Dugan was the first female director of the Defense Advanced Research Projects Agency (DARPA), the research wing of the military that helped to develop the Internet and the Global Positioning System. Her priorities included machine intelligence, flexible and manoeuvrable shape-shifting devices, and soldier telepathy using brain-wave mapping for communicating without speech. Dugan has since run technology programmes at Google and Facebook. From 2008 to 2012, Lisa Porter was founding director of 'DARPA for spies' — the Intelligence Advanced Research Projects Activity. She led investigations into quantum computing, biometric fingerprint identification and cloaking devices.

Although there is a predominance of white women in these positions, women of colour are making inroads. Quiescence Phillips, an African American, has been selected as deputy chief information security officer of cyber command for the City of New York. And Indian computer scientist Aanchal Gupta is director of security at Facebook.

Other women are taking more underground roles — as cyberspies⁶. Former judge Shannen Rossmiller gathered global online intelligence for the FBI after the terrorist attacks of 11 September 2001. By posing as male militants from Iraq and Afghanistan in extremist chat rooms, she exposed weapon caches, bomb plots and terror cells in more than 200 operations. Indeed, 35% of people who work in the intelligence field are female.

Meanwhile, cybersleuths like Kimberly Ritter are fighting crimes such as online sex trafficking. Ritter's work inspired computer scientists to develop an app called Traffickcam that enables the public to upload photographs of hotel rooms to a database. By matching these to images in advertisements for escorts, law enforcement agencies can improve the way they track down victims and their traffickers.

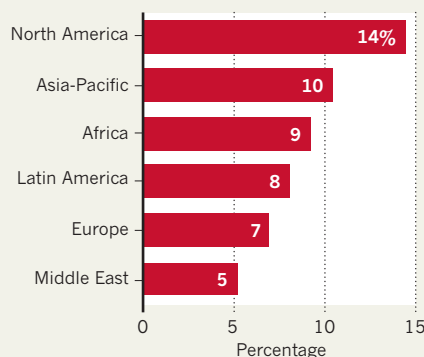
But breaking the glass ceiling is not enough. Institutional barriers also need to be overcome.

Recognize diverse expertise. Despite being few in number, female candidates for cybersecurity jobs tend to be more educated

WOMEN IN CYBERSECURITY

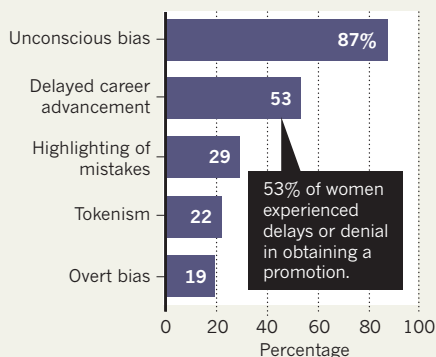
Hostile work cultures, stereotypes and discrimination are limiting the numbers* of female cybersecurity professionals (left). Just over half of women in cybersecurity say that they have experienced gender discrimination (right), compared with 15% of men.

Female cybersecurity professionals by region

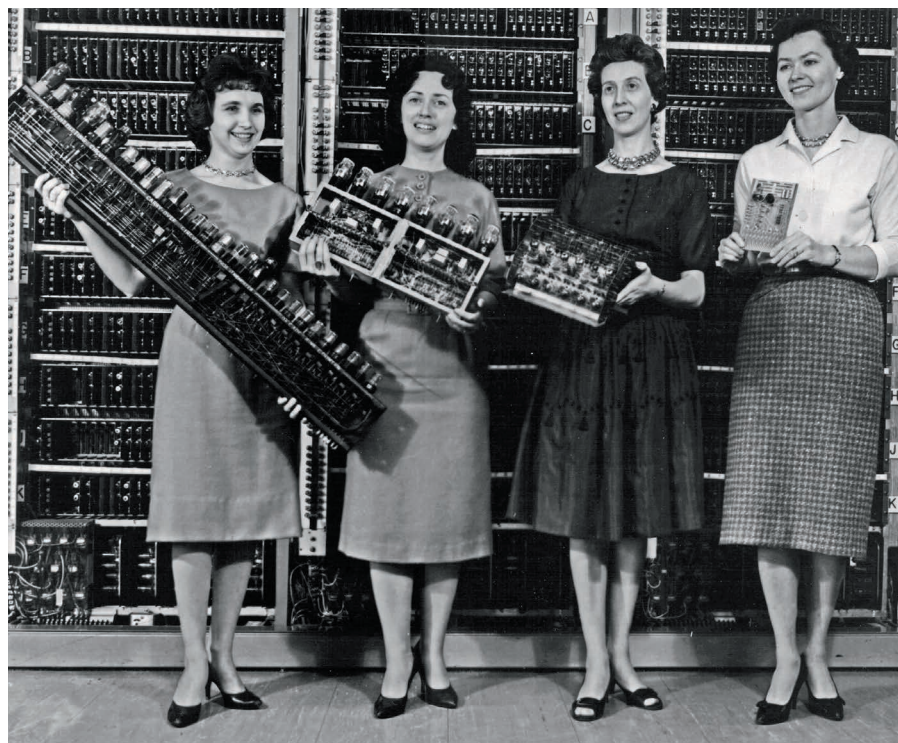


*19,641 employees sampled.

Types of workplace discrimination reported



SOURCE: 2017 GLOBAL INFORMATION SECURITY WORKFORCE STUDY



Female programmers of the Electronic Numerical Integrator And Computer (ENIAC) at the University of Pennsylvania, Philadelphia, in the 1940s.

than their male counterparts. Women professionals are more likely to have a master's degree or higher (51% for women globally, compared with 45% of men). Women also tend to bring wider expertise. Although both female and male employees train extensively in computer science, information and engineering, women's degrees are more likely to come from fields such as business, mathematics and social science (44% for women, compared with 30% for men).

This is crucial because cybersecurity jobs demand diverse skills. Professionals must understand network security, risk mitigation and information protection, and be prepared for future activities in artificial intelligence, machine learning and virtual-reality mapping. They need to manage projects, navigate legal and compliance codes, and work in sectors from health care to law enforcement⁴.

But training beyond fields in science, technology, engineering and maths (STEM) is not generally rewarded. Most cybersecurity employers view computer science or engineering backgrounds as a priority. Yet even some big tech companies are changing their outlook. In 2011, Microsoft co-founder Bill Gates disparaged liberal-arts education. Now, the company's president Brad Smith argues that such training is crucial for the future of computing, especially artificial intelligence. Google executives were shocked after they analysed their workforce data and found that STEM expertise was the least important factor associated with employees' success. Instead, qualities such as being

a good coach, problem-solver or critical thinker ranked higher.

Shed sexist images. The two fields that are most closely associated with cybersecurity — IT and the military — are plagued by cultures that are hostile towards women.

The IT industry reveres the hacker persona — a loner, typically male and white, working all night and forgoing other spheres of life. Some employees might identify with this image. Nonetheless, employers often use it as a standard for hiring and promoting, even though the workforce has diversified. As ethnomathematician Ron Eglash observed in 2002, these racial and gender premises of the hacker stereotype are being challenged by emerging groups such as “Black geeks, Asian American hipsters, and geek grrrls”⁷.

Sexist attitudes are rife in California's Silicon Valley, exemplified by the memo leaked in 2017 from a male employee of Google arguing that women are biologically ill-suited to the technology field. The ride-hailing firm Uber received more than 200 claims of sexual harassment, discrimination and inappropriate behaviours from employees, according to an investigation at its San Francisco headquarters in 2017. The US federal government has filed lawsuits against tech firms for gendered wage discrimination.

The soldier image is rooted in military legacies. Women in cybersecurity often report working with an ‘old boys club’ of former intelligence and military officers⁸. Job postings call for “ninjas” and “cyberwarriors”.

The language of cybersecurity reflects this ethos of defending networks against threats from intruders. By contrast, the concept of information security — centred on creating safe, effective systems and protecting humans who use them — describes the job better and is more widely appealing to diverse practitioners, including women.

Cybersecurity conferences are notoriously male-dominated in number and hyper-masculine in behaviour. A female attendee can be the sole woman in a group of 100 men. Two of the largest cybersecurity conferences, DEF CON and Black Hat, are held in Las Vegas, Nevada — a city long associated with the objectification of women. Women at these meetings have described experiences to scholars and online peer networks of being groped and asked by male attendees if they are a secretary or a prostitute.

The impacts are clear. More than half of women in cybersecurity around the world (51%) say that they have experienced gender discrimination, compared with 15% of men. This is sometimes overt but more often subtle — through tokenism, unconscious bias, the highlighting of mistakes, or denials and delays in career advancement. Female cybersecurity employees report a pay gap of 3–6% compared with male employees. This is much smaller than the 28% wage gap in the computing industry, but such losses accumulate over the course of a career. In the past couple of years, the gap has widened for female managers.

Women in technology leave their jobs at a higher rate than their male colleagues. The costs for US firms of losing and rehiring professionals who quit because of gender, racial, sexuality or religious bias are around \$64 billion annually.

Realize that women and girls are prime targets of cybercrime. Women in the United States were 26% more likely than men to experience identity theft in 2008, often involving the fraudulent use of a bank account or credit card. Two-thirds of victims lost money. On average, women also took longer than men to notice the breach: 83 versus 45 days. This is partly because men are more likely to bank and shop online, and so receive automatic notifications within hours — much faster than the weeks it can take to spot unauthorized transactions on monthly financial statements. Women in the United States also live in poverty to a greater degree than men do, and are less able to invest in online security services, freezing credit scores or hiring lawyers.

Some cybercrimes, such as ‘sextortion’, are directed at women and girls. Criminals trick their targets, using e-mails infected with software viruses, for instance, to gain access to their computers. Perpetrators can search for photos on hard drives or use webcams to

watch women and girls. They then blackmail targets by threatening to post images and videos on child pornography websites, to compel further sexual activity. The extortion can extend for years, because images can be permanently installed on multiple platforms. One perpetrator might have hundreds of targets. For example, 78 criminal cases of sextortion filed since 2016 may involve as many as 6,500 victims across 52 jurisdictions (US and international), 29 US states or territories and 3 foreign countries.

Strangers are usually assumed to be the culprits. In reality, women's spouses, boyfriends or family members often initiate security breaches. Abusers glean passwords through key-logging devices or by coercing their targets to hand over passwords. Misuse of technology online is linked to physical abuse offline. In a national survey, 97% of US domestic abuse shelters reported that their female clients had been harassed through technology.

NEXT STEPS

Cybersecurity needs to become more accepting of and welcoming to women. Schools and universities should emphasize the field's creativity and real-world applications. Programmes that promote technology for young women and girls should be encouraged, such as Girls Who Code, which aims to increase women in computer science and engineering. In another initiative, the US Girl Scouts organization is rolling out cybersecurity as one of its skill badges.

Media outlets and conference organizers should invite female experts and speakers. There are more than enough to choose from: the Grace Hopper Celebration (named after one of the first women coders) meets annually and brought together around 18,000 women technologists last year. WiCyS and the Diana Initiative are smaller but more specialized networks that also promote women in cybersecurity.

Employers should craft job advertisements that use inclusive wording, and seek candidates beyond computer science and the military, such as from the social sciences, humanities, law and public policy. Recruiters should use diverse selection panels



NASA mathematician Katherine Johnson did the calculations that put men on the Moon.

and gender-blinded screening of résumés. Women should be hired in cohorts to avoid isolation and experiences of tokenism.

Retaining women requires fair career trajectories. Employers should track and enforce gender parity in evaluations, promotions and salaries. Some firms are showing progress. The Anita Borg Institute's list of

"Retaining women requires fair career trajectories."

Top Companies for Women Technologists highlighted the firms Accenture, Geico and ThoughtWorks as

its winning examples in 2017.

Researchers need to answer pressing questions about gender in cybersecurity. For example, how can retail and credit firms better protect women from the fallout of hacks and identity theft? How can domestic partners be prevented from accessing passwords? Regional disparities also need examining. For instance, Europe has fewer women in cybersecurity (7%) than North America does (14%), despite similar levels of industrialization. India has a better record of graduating female computer programmers (30%) than does the United States (21%)⁹.

Cybersecurity professionals and organizations need to build partnerships to understand the types of people who are affected by cybercrime. In the United States, organizations might include the National Network to End Domestic Violence, the National Association for the Advancement of Colored People, UnidosUS (a non-profit Latino advocacy organization), the Federal Trade Commission and the Consumer Financial Protection Agency.

The purpose and practice of cybersecurity must be questioned. Huge amounts of money,

technology and resources are often allocated to command-and-control strategies that can be inefficient and counterproductive¹⁰. Common-sense solutions, such as enforcing implementation of security patches and strengthening privacy protections on consumer electronics, can be simpler, cheaper and more effective.

Security systems must protect everyone, equally. Celebrating, attracting, training and retaining a greater diversity of scholars and technicians in information-security research, business and governance are steps in that direction. ■

Winifred R. Poster is a lecturer in international affairs at Washington University, St. Louis, Missouri, USA. e-mail: wposter@wustl.edu

1. Browne, S. *Dark Matters: On the Surveillance of Blackness* (Duke Univ. Press, 2015).
2. Fagone, J. *The Woman Who Smashed Codes* (HarperCollins, 2017).
3. Light, J. S. *Technol. Cult.* **40**, 455–483 (1999).
4. Shumba, R. et al. Cybersecurity, Women and Minorities. *Proc. ITICSE-WGR June, July*, 1–14 (2013).
5. Shetterly, M. L. *Hidden Figures* (HarperCollins, 2016).
6. Poster, W. R. in *Globalization, Technology Diffusion and Gender Disparity: Social Impacts of ICTs* (eds Pande, R. & van der Weide, T. P. 247–260 (IGI Global, 2012).
7. Eglash, R. *Social Text* **20**, 49–64 (2002).
8. D'Hondt, K. *Women in Cybersecurity*. Master's thesis, Harvard Kennedy School (2016).
9. Poster, W. R. *Gen. Sex. Fem.* **1**, 37–52 (2013).
10. Molotch, H. *Against Security* (Princeton Univ. Press, 2014).

CORRECTIONS

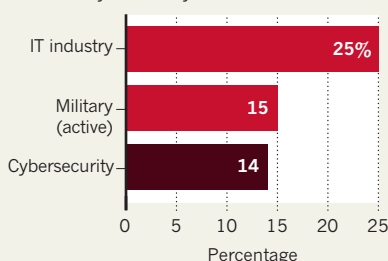
In data used for the graph 'Open countries have impact' in the Comment 'Open countries have strong science' (*Nature* **550**, 32–33; 2017), some articles classified into more than one field were counted twice. The online version of the graph has been updated to show disaggregated counts. The categorization of countries has not changed.

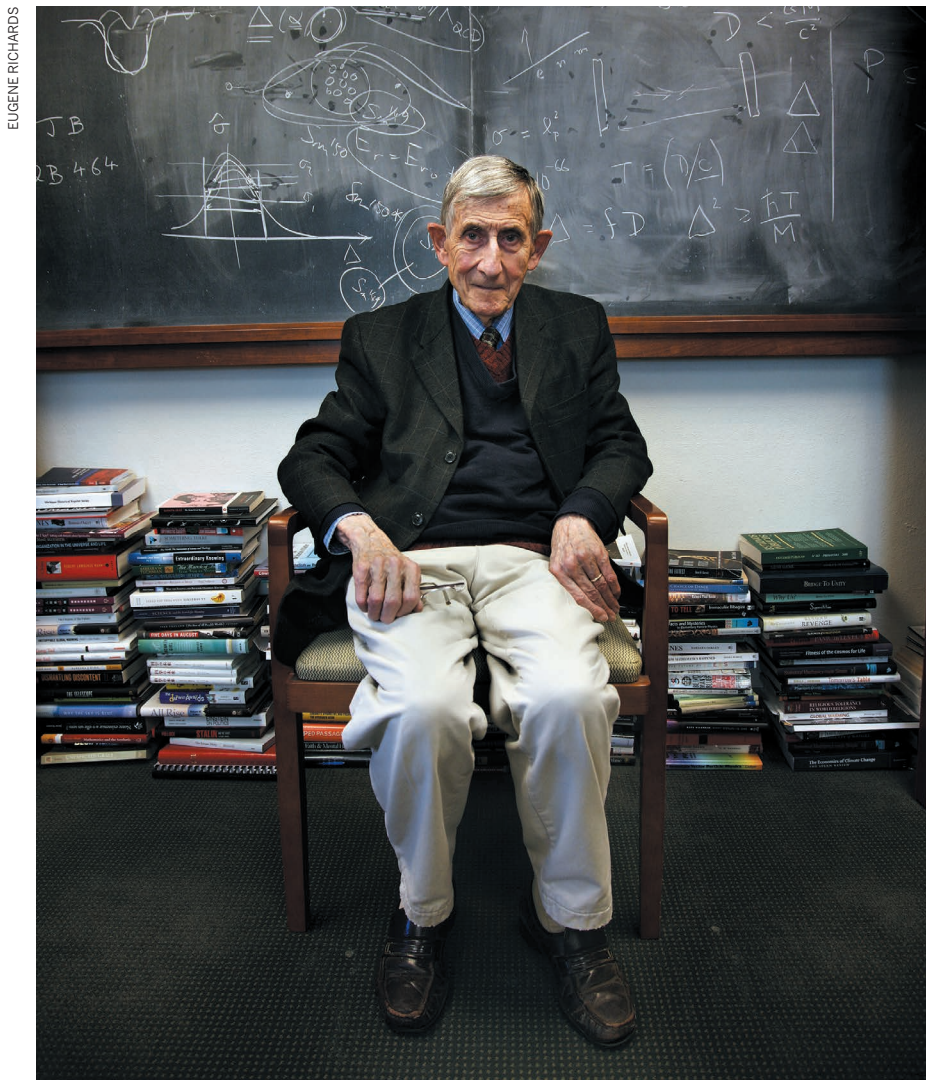
In the Comment 'How to make replication the norm' (*Nature* **554**, 417–419; 2018) mistakenly stated that only authors of non-replicated articles from the 3ie project described antagonism. In fact, authors of both replicated and non-replicated studies did. Also, the number of replications was 21, not 20. The data set is now available as supplementary information.

The Comment 'Smartphones are bad for some teens, not all' (*Nature* **554**, 432–434; 2018) misrepresented the results from the study on depressive episodes. The initial year should have been 2005, not 2004. And the rise between 2005 and 2014 should have been in percentage points, not per cent.

SISTER FIELDS

Related fields have relatively more women than does cybersecurity.





Physicist Freeman Dyson, pictured in 2009.

PHYSICS

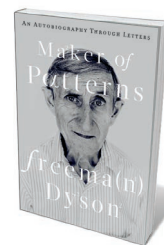
A life of scientific delight

Ann Finkbeiner is charmed by the originality and acuity in Freeman Dyson's letters.

Freeman Dyson is a mathematically inclined physicist who proved in 1949 that competing theories of quantum behaviour were equivalent. In a career spanning seven decades, he branched out into myriad fields. They included condensed-matter physics, nuclear reactors, astronomical technology, extraterrestrial habitation and advising the US government on national security, sometimes as part of the

elite post-war research group JASON.

Dyson grew up in Winchester, UK, and left in 1941 for the University of Cambridge, where he studied mathematics. There, he began a lifelong habit of writing regular letters home. Now, aged 94 and living in the United States, he has published some of them in *Maker of Patterns*. They cover a remarkable range of scientific interests, acquaintances, opinions and adventures.



Maker of Patterns:
An Autobiography
Through Letters
FREEMAN DYSON
Liveright: 2018.

The patterns Dyson says he made were first those of ideas in mathematics and physics, and then those in his writing about literature and history. Readers might hope that Dyson's own pattern — the reference frame in which his remarkable range becomes a coherent whole — would be found in his letters.

As a writer who has interviewed Dyson, I would advise against such hope.

The letters read like a travel journal written for people he loves and trusts. "I think the reason I write so openly is just this," he tells his parents in 1949, "that all these adventures in this strange new world are still somewhat unreal to me, and in writing to you about them, I bring them in contact with my familiar world and lend them some of your reality."

Dyson notices everything. He describes Americans' friendliness as a result of their inattention to the past, and thus of their loneliness in time. He watches astronomers react to an unexpected discovery: how they were impressed into silence and how the theorists mumbled but could say nothing coherent. He quotes and describes children, his own and other people's: "the children have come out of doors in summer clothes ... [The] grass is green and the crocuses are out. The children are more brightly coloured than the crocuses."

He describes his "extraordinary collection of friends". At Cambridge's Trinity College, he writes, he could always find its splendid mathematicians — including G. H. Hardy — at the local billiard table. At Cornell University in Ithaca, New York, he worked with physicists Hans Bethe ("he says he would have been willing to use atomic bombs to defeat Germany, so convinced was he of the evil character of the government") and Richard Feynman ("sizzling with new ideas, most of which are more spectacular than helpful"). At the Institute for Advanced Study in Princeton, New Jersey, he had a complex friendship with atom-bomb pioneer J. Robert Oppenheimer ("he is moving around nervously all the time, never stops smoking, and I believe that his impatience is largely beyond his control").

Some of Dyson's most vivid descriptions, however, are about encounters and individuals outside science. A theology student at Cambridge "has a deep and resonant bass voice and talks in a slow and rhythmical way, with a solemn ring on the open vowels which is enough to send shivers down one's spine". A disabled boat-builder and his girlfriend "walk slowly down the beach ... like the last act of *King Lear*, when the crazy old king and his faithful daughter Cordelia ►

► are led away to their doom”.

He says what you wouldn't expect; if Dyson has a pattern, perhaps it is contrariety. He prefers the students at Haverford College, a liberal-arts university in Pennsylvania — who are “ostentatiously ill-dressed, uncombed, and unwashed” — to Princeton's “pampered” ones. He contradicts himself. His wartime job at the UK Royal Air Force's Bomber Command, analysing the effect of Allied firebombing on German cities, left him with a “permanently bad conscience”. Yet he understood the joy of German sailors who torpedoed Allied fuel tankers because he felt “elation” when the firebombing succeeded.

“All these adventures in this strange new world are still unreal to me.”

In the late 1950s, he worked with the defence contractor General Atomic in La Jolla, California, on a project he thought would be “legendary”: a spaceship called Orion, ill-advisedly powered by nuclear explosions. Because building it required nuclear testing, he publicly opposed the Nuclear Test Ban Treaty; after Orion was defunded, he supported the treaty. Later, he became president of the anti-nuclear-proliferation Federation of American Scientists. He proposed the ‘no first use’ policy on nuclear weapons, while admiring his theoretical-physicist friend Edward Teller for standing against the treaty. Dyson's contradictions might seem confusing, but I can hear him politely and methodically explaining the rationality of seeing reality from different reference frames, of holding mutually opposing views and of being free to change one's mind.

The one Dysonian pattern for which the letters hold unequivocal evidence is delight. He uses the word often and invokes it even more: “Today I discovered a little theorem which gave me some intense moments of pleasure. It is beautiful and fell into my hand like a jewel from the sky.” In the book's last letter, he writes to his sister about being robbed and hit on the head. Preparing to die, he notices the sunlight on the bushes and thinks, “life has been good to me and this death is also good, with the bright sun and the green bushes”.

Maybe with some people, you don't look for patterns. You just enjoy their multivariate company. ■

Ann Finkbeiner is a freelance science writer in Baltimore, Maryland, and author of *The Jaxons*. She blogs at www.lastwordonnothing.com. e-mail: anniefk@gmail.com



Niels Bohr (left) with Albert Einstein in the late 1920s, when quantum mechanics was in its infancy.

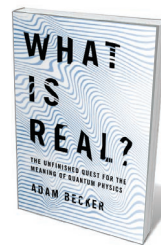
PHYSICS

Quantum-theory wars

Ramin Skibba explores a history of unresolved questions beyond the Copenhagen interpretation.

All hell broke loose in physics some 90 years ago. Quantum theory emerged — partly in heated clashes between Albert Einstein and Niels Bohr. It posed a challenge to the very nature of science, and arguably continues to do so, by severely straining the relationship between theory and the nature of reality. Adam Becker, a science writer and astrophysicist, explores this tangled tale in *What Is Real?*.

Becker questions the hegemony of the Copenhagen interpretation of quantum mechanics. Propounded by Bohr and Werner Heisenberg in the 1920s, this theory holds that physical systems have only probabilities, rather than specific properties, until they're measured. Becker argues that trying to parse how this interpretation reflects the world we live in is an exercise in opacity. Showing that the evolution of science is affected by historical events — including sociological,



What Is Real?: The Unfinished Quest for the Meaning of Quantum Physics
ADAM BECKER
Basic: 2018.

cultural, political and economic factors — he explores alternative explanations. Had events played out differently in the 1920s, he asserts, our view of physics might be very different.

Becker lingers on the 1927 Solvay Conference in Brussels, where 29 brilliant scientists gathered to discuss the fledgling quantum

theory. Here, the disagreements between Bohr, Einstein and others, including Erwin Schrödinger and Louis de Broglie, came to a head. Whereas Bohr proposed that entities (such as electrons) had only probabilities if

EMILIO SEGRE VISUAL ARCHIVES/AIP/SPL

they weren't observed, Einstein argued that they had independent reality, prompting his famous claim that "God does not play dice". Years later, he added a gloss: "What we call science has the sole purpose of determining what is." Suddenly, scientific realism — the idea that confirmed scientific theories roughly reflect reality — was at stake.

Quantum phenomena were phenomenally baffling to many. First was wave-particle duality, in which light can act as particles and particles such as electrons interfere like light waves. According to Bohr, a system behaves as a wave or a particle depending on context, but you cannot predict which it will do.

Second, Heisenberg showed that uncertainty, for instance about a particle's position and momentum, is hard-wired into physics. Third, Bohr argued that we could have only probabilistic knowledge of a system: in Schrödinger's thought experiment, a cat in a box is both dead and alive until it is seen. Fourth, particles can become entangled. For example, two particles might have opposite spins, no matter how far apart they are: if you measure one to be spin up, you instantly know that the other is spin down. (Einstein called this "spooky action at a distance".)

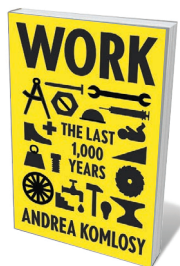
Becker explains how these observations challenge locality, causality and determinism. In the classical world of billiard balls, projectiles and apples falling from trees, they were never problems.

Sifting through the history, Becker shows how Bohr, as an anti-realist, brought to his side many rising physicists, including Heisenberg, Wolfgang Pauli and Max Born. Einstein, however, persistently argued that the Copenhagen interpretation was incomplete. He conjectured that there might be hidden variables or processes underlying quantum phenomena; or perhaps 'pilot waves', proposed by de Broglie, govern the behaviour of particles. In 1932, mathematician John von Neumann produced a proof that there could be no hidden variables in quantum mechanics. Although mathematically correct, it was revealed to be flawed decades later. But the damage had been done: the potentially viable alternatives conceived by Einstein and de Broglie remained relatively unexplored. The Copenhagen interpretation had taken hold by the 1930s, and textbooks today state that Bohr's view 'won'.

Thus, the Solvay Conference can be seen as a stand-off between two mathematically equivalent but fundamentally different paradigms: Bohr's instrumentalist view of quantum physics and Einstein's realist one. In science, a dominant paradigm determines which experiments are done, how they're interpreted and what kind of path a research programme follows.

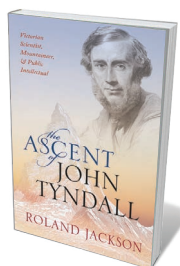
But what if a field picks the wrong paradigm? Becker shows how, in the 1950s and 1960s, a handful of physicists dusted off the ▶

Books in brief



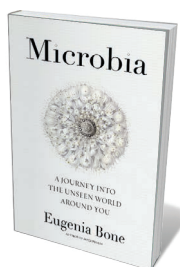
Work: The Last 1,000 Years

Andrea Komlosy, trans. Jacob K. Watson & Loren Balhorn VERSO (2018)
Employment outside the home became the globally dominant form of work only after 1900. Yet in policy and discourse, it has eclipsed older, more informal modes of production. Social historian Andrea Komlosy probes the joins between them in this sweeping chronicle, from 1250 — the dawn of globalization and urbanization — to today, when international corporations threaten labour standards. This is a book teeming with insights, from the contempt for manual labour in ancient Greece to the historical tendency for all kinds of subsistence tasks to be "housewife-ized" into unpaid domestic labour.



The Ascent of John Tyndall

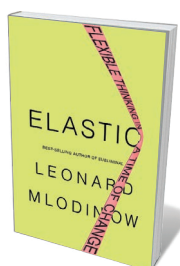
Roland Jackson OXFORD UNIVERSITY PRESS (2018)
A leading light in climate science, germ theory and magnetism, Victorian researcher John Tyndall seemed a man for all sciences. He discovered the greenhouse effect and why the sky is blue, and as a public intellectual, hobnobbed with physicists Michael Faraday and Hermann von Helmholtz, and poet Alfred Tennyson. This splendid monument of a biography by Roland Jackson tracks Tyndall's rise from rural Ireland to laboratorial glory days in Britain and Europe. The experimentalist's obscurity now, Jackson avers, may be down in part to his death in 1893: he just missed the era of relativity and the quantum.



Microbia

Eugenia Bone RODALE (2018)

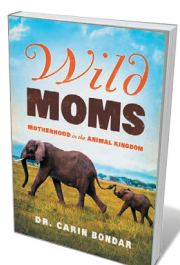
At the age of 55, journalist Eugenia Bone returned to university to take a microbiology course, reasoning that "humility is the entry point for studying nature". The microbial matrix of life came into focus. Bone's is an exquisitely observant, often amusing voyage from the origins of life to antibiotic resistance, punctuated by scientific revelations. (The compound geosmin, produced by *Streptomyces* bacteria, gives soil its sweet smell; bacteria nucleate rain; a quadrillion microbes pack part of a cow's stomach.) The devil's in the detail, Bone finds, but what enchants is the connectivity of this 'hidden' ocean.



Elastic: Flexible Thinking in a Constantly Changing World

Leonard Mlodinow PANTHEON (2018)

Physicist Leonard Mlodinow revels in the neuroscientific. Here he extends his explorations with an in-depth study of agile creative thinking in a hectic age. Among the psychological factors Mlodinow isolates are a yen for novelty, and the capacity to reconcile diverse ideas. He plunges deep into the human brain, marshalling compelling research both on brain basics and on outlier issues such as mental blocks. Perhaps most gripping is his take on the "default network" — the brain structures that govern richly creative activities such as daydreaming, the "dialogue we have with ourselves".



Wild Moms

Carin Bondar PEGASUS (2018)

Biologist and broadcaster Carin Bondar's tour of the vagaries of motherhood in the animal kingdom is a thrills-and-spills survey, from brood parasitism to cooperative breeding. The diversity of evolutionary solutions to maternity is mind-bending. We learn, for instance, how baby koalas ingest their mothers' faeces to inoculate their gut microbiomes, and that female gastric-brooding frogs vomited up their froglets. Ultimately, however, this is less a synthesized narrative than a biological litany. [Barbara Kiser](#)

► theories of Einstein and de Broglie and turned them into a fully fledged interpretation capable of shaking up the status quo. David Bohm argued that particles in quantum systems existed whether observed or not, and that they have predictable positions and motions determined by pilot waves. John Bell then showed that Einstein's concerns about locality and incompleteness in the Copenhagen interpretation were valid. It was he who refuted von Neumann's proof by revealing that it ruled out only a narrow class of hidden-variables theories.

The scientific community greeted Bohm's ideas coolly. A former mentor, J. Robert Oppenheimer, said: "if we cannot disprove Bohm, then we must agree to ignore him". And, as Becker shows, Bohm's leftist views led to an appearance before the House Un-American Activities Committee, and subsequent ostracization.

Bohm's contemporary, physicist Hugh Everett, delivered another challenge to the Copenhagen interpretation. In 1957, Everett set out to resolve the 'measurement problem' in quantum theory — the contradiction between the probabilistic nature of particles at the quantum level and their 'collapse', when measured, into one state at the macroscopic level.

Everett's many-worlds interpretation posited no collapse. Instead, probabilities bifurcate at the moment of measurement into parallel universes — such as one in which Schrödinger's cat is alive and another in which it's dead. Although an infinite number of untestable universes seems unscientific to some, many physicists today view the theory as important.

The book has a few minor shortcomings. Becker gives too much space to recent applications building on Bell's research, and too little to new developments in the philosophy of science. Yet he, like cosmologist Sean Carroll in his 2016 *The Big Picture* (R. P. Crease *Nature* 533, 34; 2016), does make an explicit case for the importance of philosophy. That's a key call, with influential scientists such as Neil deGrasse Tyson dismissing the discipline as a waste of time.

What Is Real? is an argument for keeping an open mind. Becker reminds us that we need humility as we investigate the myriad interpretations and narratives that explain the same data. ■

Ramin Skibba is an astrophysicist turned science writer based in San Diego, California. e-mail: raminskibba@gmail.com

"Becker reminds us that we need humility as we investigate the myriad interpretations that explain the same data."



A Pan Am shuttle prepares to dock at the international Space Station V in this classic scene from the film.

IN RETROSPECT

2001: A Space Odyssey

Fifty years on, the masterful science-fiction film looks more prophetic than ever, reflects **Piers Bizony**.

In 1968, film-maker Stanley Kubrick and his screenwriting colleague, science-fiction author Arthur C. Clarke, presented *2001: A Space Odyssey*. Half a century later, this unprecedentedly detailed speculation about our place in the cosmos and our evolving relationship with technology is regarded as one of the great landmarks in cinema.

The influence of *2001* on the design of subsequent space-film hardware and special effects has been pervasive. However, in terms of artistic and philosophical bravura, it has been a harder act to follow. In 2007, director Ridley Scott (of *Blade Runner* and *Alien* fame) told a Venice Film

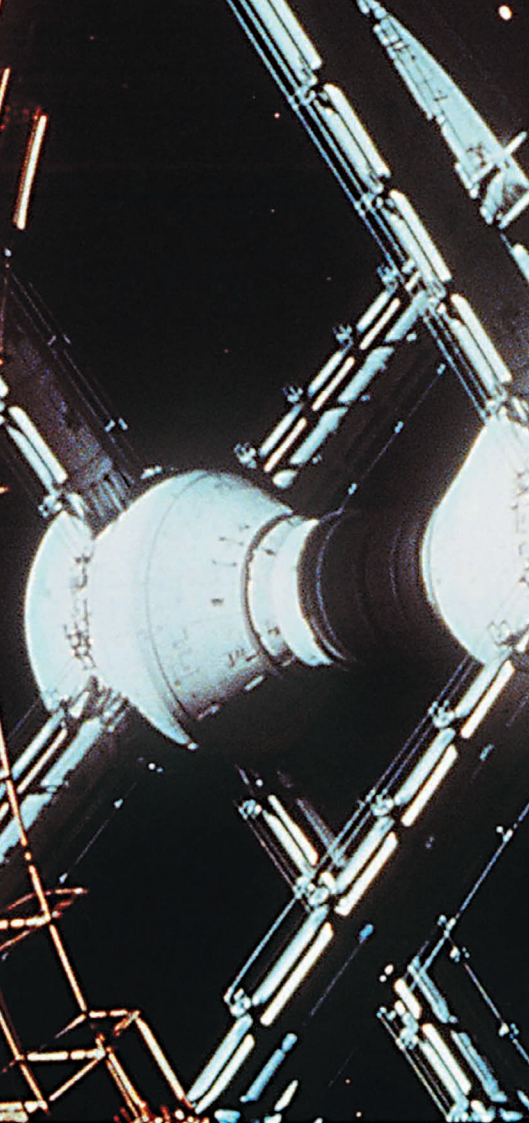
2001: A Space Odyssey

DIRECTOR: STANLEY KUBRICK; CO-WRITER: ARTHUR C. CLARKE
Metro-Goldwyn-Mayer: 1968.

Festival audience: "After *2001*, science fiction is dead."

The narrative was ambitious, to say the least. Prehistoric apemen have a mind-altering encounter with an alien black monolith. Four million years later, a giant spacecraft is sent to Jupiter on a mysterious mission. On board are two astronauts, three hibernating scientists and a seemingly sentient computer, HAL 9000. Hovering above Jupiter, another monolith waits.

Monoliths aside, *2001* was prescient in almost all its detailed predictions of



BOTH PICTURES: WMGM/STANLEY KUBRICK PRODUCTIONS/KOBAL/REX/SHUTTERSTOCK

twenty-first-century technology. For instance, in August 2011, the Samsung electronics group began a defence against a claim of patent infringement by Apple. Who invented the tablet computer? Apple claimed unique status for its iPad; Samsung presented a frame from *2001*.

Samsung noted that the design claimed by Apple had many features in common with that of the tablet shown in the film clip — most notably, a rectangular shape with a display screen, narrow borders, a flat front and a thin form. In an era when computers still needed large rooms to accommodate them, Kubrick's special-effects team rigged hidden projectors to enliven devices that looked as though you could hold them in one hand. Only the need to trim the film's running length prevented ingenious mock-ups of touch-sensitive gaming screens and electronic newspapers from making the final cut.

Indeed, *2001* got much right about the twenty-first century, including the psychological costs of our reliance on technology. Kubrick's astronauts look listless inside vehicles perfectly capable of functioning without them. Dialogue in the film is deliberately banal. HAL gets all the best lines, even as he kills off most of his human

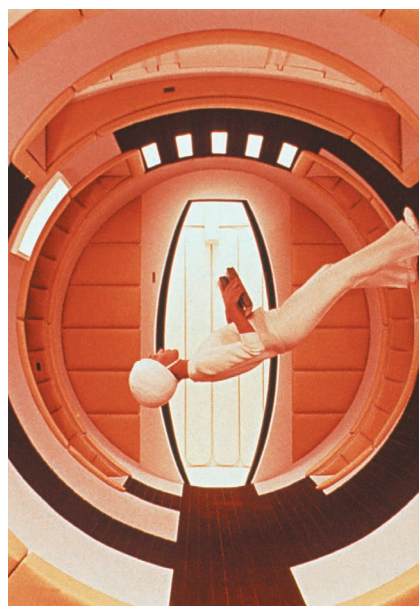
companions on board the Jupiter-bound spaceship. As our machines smarten up, we will dumb down, Kubrick suggested. The crass human discourse often found on today's algorithmically clever social-media platforms seems to bear out his pessimism.

Certainly, in the film, the surviving astronaut's final conflict with HAL prefigures a critical problem with today's artificial-intelligence (AI) systems. How do we optimize them to deliver good outcomes? HAL thinks that the mission to Jupiter is more important than the safety of the spaceship's crew. Why did no one program that idea out of him? Now, we face similar questions about the automated editorship of our searches and news feeds, and the increasing presence of AI inside semi-autonomous weapons.

We also face the startling fact that 1960s assumptions about the progress of AI were optimistic (computing pioneer Marvin Minsky served as a consultant for 2001). No machine available today can match HAL's performance. Just look at the YouTube spoofs of "digital personal assistants" reimagined as HAL: Kubrick's red-eyed genius outsmarts them all.

Where the film made technical errors, it did so in ways that were flawed rather than merely wrong. Kubrick's machinery is a matter for nerdy analysis instead of the ridicule usually reserved for outdated fantasies. Rocket experts happily argue about the film's docking arrangements and control-panel layouts; but none of the machinery is out of bounds in terms of what we could actually achieve.

And Kubrick and Clarke confidently predicted space projects that still lie many years, or decades ahead, 17 years after the film's eponymous dateline. Famously,



A shuttle flight attendant walks in zero gravity.

Kubrick gave us the most persuasive space station in all of science fiction, a gigantic twin-wheeled "Orbiter Hilton". It was complete with coffee-vending machines, garish designer chairs in the lounges, phone booths accepting credit cards, and panoramic windows offering spectacular views of the Earth (something so familiar to the film's spacefarers that they ignore it).

Space planners of the 1960s assumed that artificial gravity would be essential. In the event, the International Space Station was designed for science experiments that depend on cancelling out gravity's influence, not replicating it. Now the wheel has turned full circle, as we recognize the debilitating effects of long-term weightlessness on the human body. Aerospace engineers are starting to think about how we could make revolving habitats for future deep-space missions, or orbiting hotels for private adventurers.

As for such doughty explorers, recent triumphs of Elon Musk's SpaceX rocket company validate Kubrick's use of Pan Am as his space-shuttle carrier of choice. Although Pan Am no longer exists, the corporate conquest of the cosmos is under way, just as *2001* predicted. For today's orbital entrepreneurs, the film is not just a beautiful piece of science fiction. It is a technical manifesto — an inspirational call to arms, around which a real commercial space industry is being founded.

Of course, *2001* also suggests some kind of non-terrestrial influence over human progress. After unsatisfactory experiments trying to depict aliens, Kubrick chose only to hint at their presence, allowing those black slabs to stand in for — well, for what, exactly?

Today, we know that planetary systems orbiting other suns are the rule rather than the exception, and that prebiotic molecules suffuse the vast, interstellar clouds of dust and gas from which new stars and planets are born. Have any of these molecules sparked into life on other worlds, perhaps giving rise to intelligent entities? If so, might some of them be more advanced than we are?

Should we watch out for superior "aliens" closer to home, and guard against AI systems one day supplanting us in the evolutionary story yet to unfold? Or does the absence of anything like HAL, even after 50 years, suggest that there is, after all, something fundamental about intelligence that is impossible to replicate inside a machine? Until we know the answers to such profound questions, *2001: A Space Odyssey* cannot stale. ■

Piers Bizony is a science journalist, space historian and author.
e-mail: piersbizony@icloud.com

Correspondence

Climate and conflict: risk multipliers

To our knowledge, no one in the field of climate research has suggested that climate change could be the “sole cause” of war, violence, unrest or migration (see *Nature* **554**, 275–276; 2018). We argue that viewing climate change instead as a risk multiplier, influencer or co-factor can help to inform rather than inflame this important discussion.

In this way of thinking, environmental and ecological factors interact with social determinants, including those that are economic, demographic and political, to produce phenomena such as migration, conflict and famine (D. C. Bowles *et al. J. R. Soc. Med.* **108**, 390–395; 2015). An example you cite found that drought across sub-Saharan Africa in 1990–2011 contributed to an increased risk of rioting because it affected the region’s rain-fed agriculture (C. Almer *et al. J. Environ. Econ. Manage.* **86**, 193–209; 2017). This type of agriculture was thus a contributory factor to the likelihood of rioting — a risk that was then magnified by drought.

Including such environmental factors and multipliers will improve understanding of the causes of conflict. Without these, models for studying conflict could become an oversimplified form of social determinism.

Colin D. Butler *Australian National University, Canberra, Australia.*

Ben J. Kefford *University of Canberra, Canberra, Australia.*
colin.butler1955@gmail.com

Climate and conflict: don’t oversimplify

Scientific and methodological challenges make it hard to evaluate links between climate change and human conflict. We find the criticisms made in a review of these challenges (see C. Adams *et al. Nature Clim. Change* **8**, 200–203; 2018), as

outlined in your Editorial (*Nature* **554**, 275–276; 2018), to be a flawed oversimplification.

One criticism is that the field suffers from a potential sampling bias because research focuses only on regions of violent conflict to draw conclusions about the social and political effects of climate change. In fact, researchers use such cases to identify the variety of factors that produced these conflicts, just as epidemiologists study disease outbreaks to understand the factors that produce epidemics.

Researchers are also accused of ignoring instances of cooperation in favour of conflict. That is not why studying cooperation is important. Comparing triggers of cooperation and conflict helps to understand human responses to environmental stresses.

We agree that research results should not be overgeneralized or inappropriately used to justify causality between climate and conflict. Both practices are rare in our experience, contrary to your implication. Most researchers take pains to describe climate and weather merely as ‘contributing factors’ to some conflicts (see, for example, C. P. Kelley *et al. Proc. Natl Acad. Sci. USA* **112**, 3241–3246; 2015).

Peter H. Gleick *Pacific Institute, Oakland, California, USA.*

Stephan Lewandowsky *University of Bristol, UK.*

Colin Kelley *Center for Climate and Security, Washington DC, USA.*
pgleick@pacinst.org

Climate and conflict: no stigma

Your Editorial on sampling bias in studies linking climate change with civil unrest (*Nature* **554**, 275–276; 2018) is based on an analysis that in our view provides no evidence for biased results (C. Adams *et al. Nature Clim. Change* **8**, 200–203; 2018).

We disagree with your contention that it is “undesirable” to study risk factors for populations with a high likelihood

of conflict on the grounds that it could “stigmatize” these regions as politically unstable. The same logic would argue against studying risk factors for people who have a high chance of developing cancer for fear of stigmatizing patients. In our view, such recommendations could create bias in the literature by inhibiting research.

Studies of connections between climate and conflict should instead be motivated to identify causes of human suffering so that it can be alleviated (see, for example, M. Burke *et al. Annu. Rev. Econ.* **7**, 577–617; 2015). We do not believe that shying away from investigations in this field is an effective path towards this goal.

Solomon Hsiang *University of California, Berkeley, USA.*

Marshall Burke *Stanford University, California, USA.*
shsiang@berkeley.edu

Details matter for lab kit contaminants

We wish to clarify some points related to our reporting of facultative pathogenic bacteria in a commercially available genetic-engineering kit last year (see *Nature* **552**, 291; 2017).

Enterococcus spp., *Klebsiella pneumoniae* and *Enterobacter* spp. are among the ten microbes most frequently associated with infections in the acute-care wards of European hospitals. For users of the contaminated kits, the risk of infection through broken skin or mucosal contact was considered ‘low’ by the European Centre for Disease Prevention and Control (ECDC) — not ‘very low’, as you wrote. These bacteria were multidrug-resistant and so needed special measures for safe handling.

Regarding your implication that the Bavarian authorities withheld data on the kits, we informed all key people involved, as well as relevant institutions such as the World Health Organization and the ECDC,

about the methods and outcome of our analyses. However, the official results were not made publicly available.

We stress that the regulatory status of biological agents in Germany is unambiguous. In this case, the bacteria were classified as pathogens under the German Protection Against Infection Act and so their import, export, storage and handling needed official permission. The German Genetic Engineering Act allows genetic engineering operations only in approved installations.

Nina Koehler, Armin Baiker, Ulrich Busch *Bavarian Health and Food Safety Authority (LGL), Oberschleißheim, Germany.*
ulrich.busch@lgl.bayern.de
Competing non-financial interests declared; see go.nature.com/2hznref.

Boost children’s digital intelligence

A survey last year by my organization of 38,000 children across 29 countries (see go.nature.com/2fgrnnp), revealed that more than 50% of 8–12-year-olds were exposed to at least one cyber-related threat such as technology addiction, cyber-bullying or identity theft (see also C. Odgers *Nature* **554**, 432–434; 2018). One solution is to teach children how to develop digital intelligence (DQ) — a set of social, emotional and cognitive capabilities that helps them to minimize the risks and pressures of digital media and to maximize its opportunities.

At Singapore’s award-winning DQ Institute, we work with governments and international partners such as the World Economic Forum and the Organisation for Economic Co-operation and Development to improve digital education, culture and innovation through cross-sector collaboration, global dialogue and big data research (www.dqinstitute.org).

Yuhyun Park *Nanyang Technological University, Singapore.*
park@dqinstitute.org

John Sulston

(1942–2018)

Nobel-prizewinning champion of the Human Genome Project and open data.

The principle that genomic data should be universally shared without commercial involvement owes its widespread acceptance largely to John Sulston. As leader of the British contribution to the international Human Genome Project, Sulston persuaded funders and colleagues of the crucial importance of making a complete, high-quality sequence freely available to the global scientific community. His commitment stemmed from a moral certainty that profit as a motive had no place in science.

In 2002, he shared the Nobel Prize in Physiology or Medicine for his contribution to understanding how genes control the fate of cells in the developing roundworm *Caenorhabditis elegans*. In his work on the worm cell lineage and, later, genome sequencing, Sulston promoted the idea that investing in large-scale data collection without a specific hypothesis has long-term benefits.

Sulston, who died on 6 March, was the son of an English clergyman and a teacher. He took to heart his parents' indifference to material wealth and their desire to work for the common good. Mechanics fascinated him from an early age and, by dissecting dead animals, he began to see living things, too, as machines. He won a scholarship to study natural sciences at the University of Cambridge, UK.

His PhD, also at Cambridge, was on the synthesis of oligonucleotides — the building blocks of nucleic acids such as DNA and RNA. His flair as an experimentalist secured him a position as a postdoc with organic chemist Leslie Orgel at the Salk Institute for Biological Studies in La Jolla, California, investigating the origins of life. Sulston arrived in 1966 to study the replication of nucleic acids, and, for the first time he understood the evolutionary context of his work in chemistry.

There, too, Sulston met Francis Crick, co-discover of the DNA's double helix; Crick recommended him for a post at the Medical Research Council's Laboratory of Molecular Biology back in Cambridge. In 1969, Sulston joined a small group headed by geneticist Sydney Brenner. Looking for a model organism in which to explore the interactions of genes, development and behaviour, Brenner chose *C. elegans*. The tiny worm has only 959 cells as an adult, yet it has a gut, a nervous system, gonads and a repertoire of behaviours to move, feed and reproduce.

It is also transparent. Sulston used what became his favourite tool, the Nomarski



differential-interference contrast microscope, to visualize cell nuclei in living worm larvae and, later, in the more challenging embryos. He observed and recorded the invariant sequence of cell divisions that build an adult worm. His work provided a foundation for future worm biologists to answer questions about development that have implications for other species, including our own.

Sulston spotted that certain cells are eliminated during development. He began to explore the genetics of this process with US postdoc, Robert Horvitz. On his return home to the Massachusetts Institute of Technology in Cambridge, Horvitz discovered genes that regulate programmed cell death. Later, mutations in these genes proved instrumental to understanding the uncontrolled multiplication of cancer cells. Sulston, Horvitz and Brenner shared the 2002 Nobel prize “for their discoveries concerning genetic regulation of organ development and programmed cell death”.

From 1983, Sulston set out to map and sequence the 100-million-base-pair genome of the worm. His lab worked in partnership and friendly competition with that of Robert Waterston, then at the University of Washington in St. Louis. In 1989, their collective drive motivated Jim Watson, then head of the Human Genome Project, to fund their pilot sequencing initiative. The worm project demonstrated that automated, high-throughput

sequencing of the human genome was not impossible. In 1992, the UK Wellcome Trust invited Sulston to head its new sequencing facility, the Sanger Centre (now the Sanger Institute), in Hinxton.

Leading the Sanger as part of an international consortium, Sulston had a key role in establishing the principles of data release and open access. When, in 1998, a private initiative led by Craig Venter of Celera Genomics announced its intent to sequence the human genome first, and for commercial profit, Sulston defended the principle of open data. He became, with Francis Collins — then head of the US National Human Genome Research Institute — a leading voice in persuading both the US National Institutes of Health and the Wellcome Trust to commit to the project's completion. The human genome sequence, finished to a standard of 99.99% accuracy, was published in *Nature* on 21 October 2004 (International Human Genome Sequencing Consortium *Nature* **431**, 931–945; 2004).

As Sulston predicted, the availability of the complete human genome drove research, both academic and commercial, worldwide. The work is slowly delivering clinical applications, especially in areas such as cancer, where genetic mutations determine whether a tumour will be sensitive to a therapy. Meanwhile, technology has advanced and costs dropped such that whole-genome sequencing of individuals might soon be routine.

Sulston was hands-on in the lab, personally preparing the worm-clone library for the mapping project, and decrypting the early sequencing machines so that the electronic data could be analysed directly. But he always valued collaborators with skills different from his own, and co-managed the Sanger on a board of seven people. When he asked me to co-author a book on the battle for the human genome, *The Common Thread*, he insisted that we work as equal partners.

After the human sequence was published, Sulston devoted himself to writing and speaking in support of open access and, more broadly, on the relationship between science and society. Warm and articulate, he won over audiences with his humility and passion. ■

Georgina Ferry is a science writer and biographer in Oxford, UK. She and Sulston co-authored *The Common Thread: A Story of Science, Politics, Ethics and the Human Genome* (Bantam Press, 2002).
e-mail: mgf@georginaferry.com

SCOTT BARBOUR/GETTY

Is plasticity caused by single genes?

ARISING FROM C. K. Ghalambor *et al.* *Nature* **525**, 372–375 (2015); doi:10.1038/nature15256

Whether phenotypic plasticity generally facilitates or hampers adaptive evolution is a matter of much debate. By comparing gene expression changes in guppy populations, Ghalambor *et al.*¹ suggest that ‘adaptive’ plasticity hampers adaptive evolution, whereas ‘non-adaptive’ plasticity facilitates adaptive evolution². Here, we argue that the classification of individual gene expression changes as representing either adaptive or non-adaptive plasticity is problematic. Instead of indicating adaptive evolution, the expression changes of suites of genes may be caused by potentially random changes in underlying regulatory processes. There is a Reply to this Comment by Ghalambor, C. K. *et al.* *Nature* **555**, <http://dx.doi.org/10.1038/nature25497> (2018).

Ghalambor *et al.*¹ did not directly measure organismal plasticity or evolutionary adaptation at the organismal level², but inferred evolutionary changes in plasticity from transcriptomic data (gene expression profiles). They first selected genes in the derived fish populations that are concordantly differentially expressed with respect to the ancestral population, thereby accounting for possible non-independencies in their expression data. They then continued their analysis by evaluating the selected genes one by one, thereby assuming that each gene is an independent realization of plasticity and that all convergent gene expression changes are adaptive. Yet, genes do not act in isolation to shape the organismal phenotype, but are part of regulatory networks^{3,4}. A change in the expression of one gene can propagate through the entire network, thereby affecting the expression of many downstream regulatory genes, many of which might have no consequences for the organismal phenotype. Moreover, changes in the expression of one gene strongly depend on the action of other genes^{5–7}. As a result, divergent changes at the gene expression level can be associated with convergent changes (or lack of change) at the level of organismal plasticity; whereas convergent gene expression changes do not exclude diversity in organismal response patterns. To illustrate this, we ran replicate evolutionary simulations, in which a simple gene regulatory network was selected to launch an adequate organismal response to the presence and absence of predators. In each simulation, the same optimal response readily evolved, but the network structures mediating this response differed substantially across replicate simulation runs (Fig. 1). Thus, without accounting for regulatory interactions, it is often impossible to infer the organismal consequences of gene expression changes and, hence, their adaptive value.

By stating that selection acts “more strongly to decrease plasticity in those transcripts showing the greatest non-adaptive plasticity,” Ghalambor and colleagues¹ suggest that individual genes can be classified as expressing adaptive or non-adaptive plasticity, and that the expression changes of each gene between the ancestral and derived population resulted from adaptive evolution. This gives the false impression that the observed expression changes of many genes can be regarded as a large number of independent data points, whereas a single regulatory change could generate the same expression profiles. We illustrate this by a simple model that demonstrates how a single regulatory change can yield the same results as those shown by Ghalambor *et al.*¹ without indicating adaptive evolution. To motivate this model, we need to have a closer look at their classification of adaptive and non-adaptive plasticity. In their experiment, Ghalambor *et al.*¹ transplanted Trinidadian guppies (*Poecilia reticulata*) from a habitat with predators to two predatory-free habitats. After a few generations, fish from all three locations were reared in the laboratory to examine the degree of evolutionary divergence and the plastic responses of fish

towards the presence and absence of predators. Genes in the ancestral population were classified as expressing adaptive plasticity when, in a predator-free environment, they exhibit an up- or downregulation in the direction of gene expression in the derived populations in this environment, but only if the degree of regulation is weaker in the ancestral population. Thus, plastic responses that occur in the same direction in both ancestral and derived populations but have a smaller effect size in the derived fish are considered to be non-adaptive. According to this definition, any factor that weakens the change in gene expression in the derived populations would be a source of non-adaptive plasticity. Let us assume, for example, that a large set of genes is under (partial) control of a regulatory mechanism, which affects the expression of these genes in response to different environments. An example is the hormone cortisol, which in guppies typically has higher levels in the presence than in the absence of predators⁸. Cortisol can have a major influence on gene expression⁹ and it affects organismal phenotypes in a variety of ways¹⁰. The simplest model for the action of a hormone on many genes posits that the expression level of gene *i* is given by $G_i = B_i + w_i \times H(E) + \varepsilon$, in which B_i is the baseline expression level of this gene, $H(E)$ is the environment-specific level of the hormone, and ε subsumes all stochastic effects. The effect w_i of the hormone on gene

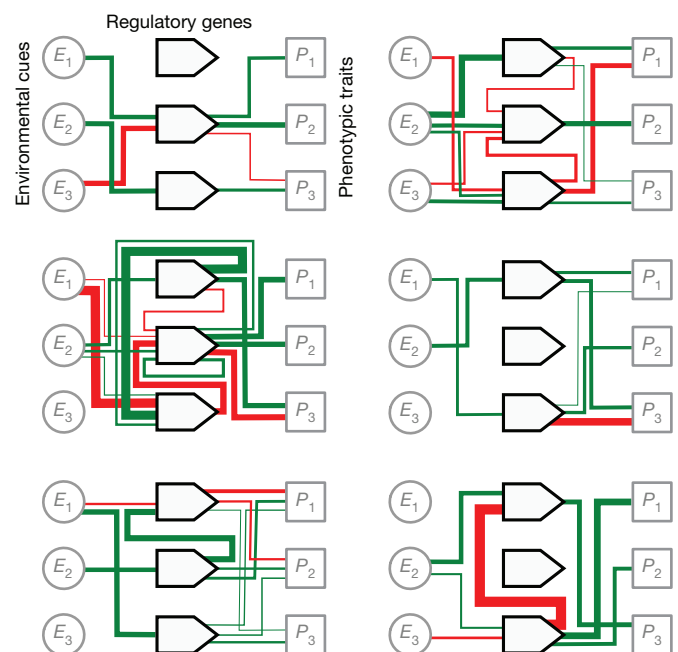


Figure 1 | Alternative gene regulatory network with identical plastic responses. Networks were selected to show an adequate phenotypic response to the presence (cues: $E_1 = 1.5$, $E_2 = 2.2$, $E_3 = 0.3$) or absence of predators (cues: $E_1 = 1$, $E_2 = 2$, $E_3 = 0.5$). The optimal response is to express three different phenotypic traits as $P_1 = P_2 = 0$, $P_3 = 1$ in the presence of predators, and as $P_1 = P_2 = 1$, $P_3 = 0$ in the absence of predators. The networks shown are the result of six replicate evolutionary simulations; all networks produce the same optimal phenotypic response, but the stimulating (green) and inhibitory (red) interactions within the networks and their strength (thickness) differ considerably. In accordance with previous models¹², we used a gene regulatory network implementation with Boolean gene expression and continuous connection weights.

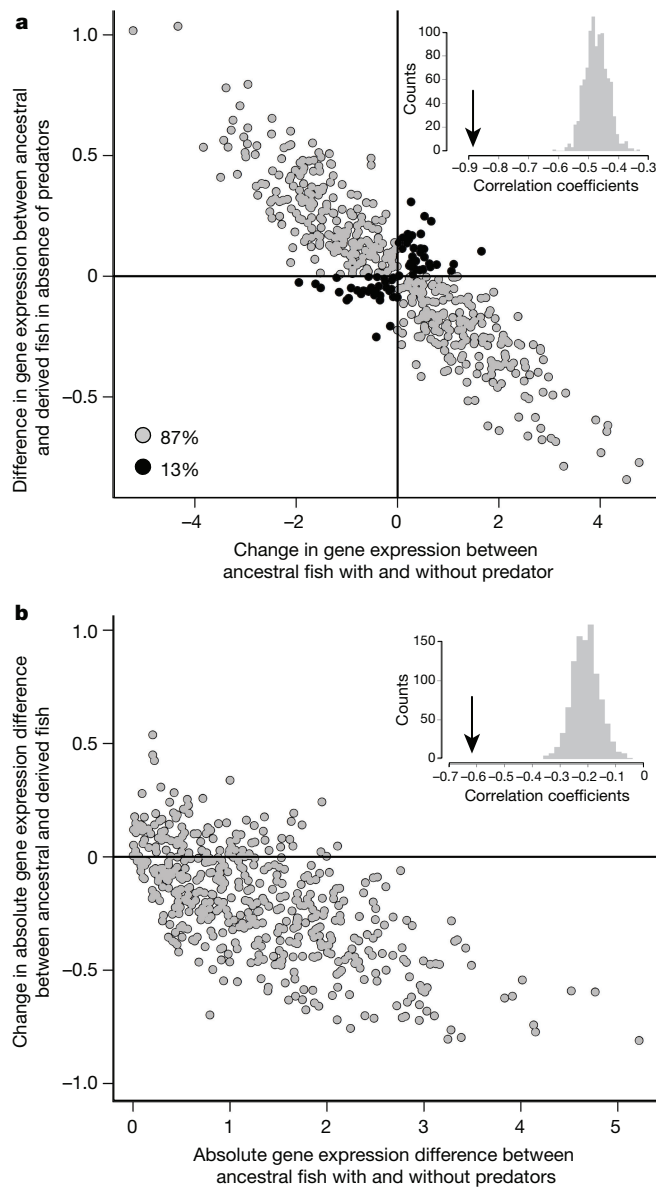


Figure 2 | Effect of small hormonal differences on gene expression patterns. a, b, In an ancestral and a derived population, the hormone-modulated expression pattern of 500 genes was assumed to follow the model $G_i = B_i + w_i \times H(E) + \varepsilon$ (see text). All parameters were the same in both populations, with the exception of hormone levels in a predator-free environment (ancestral population: $H = 1.6$; derived population: $H = 1.3$). Conducting the same analyses as in Fig. 2 and 3 of ref. 1 (panels a and b, respectively) produces very similar patterns (of high statistical significance) as those reported by Ghalambor *et al.*¹ Parameters: baseline levels (which do not affect the results) $B_i = 0.0$; $H = 0.0$ in the presence of predators; ε normally distributed with mean of 0 and s.d. of 0.1.

i can be positive or negative and large or small (in our simulation, the w_i effects are drawn from a standard normal distribution). If we now assume that fish from a derived population have a slightly lower hormone level when reared in the absence of a predator than fish from the ancestral population, this simple model reproduces the main results of Ghalambor *et al.*¹ (compare figures 2 and 3 from ref. 1 to Fig. 2a, b). Hence, a single regulatory change provides a much more parsimonious explanation for the observed transcription patterns than the large-scale

evolutionary changes postulated by Ghalambor *et al.*¹ Our alternative explanation does not refer to adaptive or non-adaptive plasticity, and it does not require evolution of gene regulation patterns (the w_i effects remain constant). It does not even require adaptive evolution of the hormone level, because the lowered hormone level of the derived populations in a predator-free environment might just reflect a random change, for example, one caused by a founder effect¹¹.

We conclude that the evolutionary analysis of transcriptomics data remains a major challenge. Ghalambor *et al.*¹ present an intriguing experiment to investigate the role of phenotypic plasticity on evolutionary adaptation, in which they show that gene expression patterns can rapidly diverge between independently evolving populations. We do not question this rapid divergence or that non-adaptive plasticity could potentially underlie this divergence, instead we want to point out that one should be cautious when inferring evolutionary changes in plasticity from transcriptomic data, especially when the underlying regulatory mechanisms are unknown. Because many loci are involved, the rapid divergence in gene expression found by Ghalambor *et al.*¹ probably reflects only a few regulatory changes, making it impossible to classify individual genes as expressing either adaptive or non-adaptive plasticity. Thus, owing to the complex relation between an organism's genotype and phenotype, it is hard to understand adaptive evolution by focusing on single genes alone.

Data Availability All data are available from the corresponding author upon reasonable request.

J. van Gestel^{1,2,3} & F. J. Weissing^{1,4}

¹Groningen Institute for Evolutionary Life Sciences, University of Groningen, 9747 AG, Groningen, The Netherlands.

email: f.j.weissing@rug.nl

²Department of Evolutionary Biology and Environmental Studies, University of Zürich, 8057 Zürich, Switzerland.

³Department of Environmental Systems Science, ETH Zürich, 8092 Zürich, Switzerland.

⁴Netherlands Institute for Advanced Study, 1012 CG Amsterdam, The Netherlands.

Received 21 October 2015; accepted 29 November 2017.

1. Ghalambor, C. K. *et al.* Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**, 372–375 (2015).
2. Merilä, J. Evolutionary biology: perplexing effects of phenotypic plasticity. *Nature* **525**, 326–327 (2015).
3. Segrè, D., Deluna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2005).
4. Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920 (2010).
5. Wagner, A. *Robustness and Evolvability in Living Systems* (Princeton Univ. Press, 2007).
6. Wagner, A. *The Origins of Evolutionary Innovations* (Oxford Univ. Press, 2011).
7. Beukeboom, L. & Perrin, N. *The Evolution of Sex Determination* (Oxford Univ. Press, 2015).
8. Fischer, E. K., Harris, R. M., Hofmann, H. A. & Hoke, K. L. Predator exposure alters stress physiology in guppies across timescales. *Horm. Behav.* **65**, 165–172 (2014).
9. Aluru, N. & Vijayan, M. M. Stress transcriptomics in fish: a role for genomic cortisol signaling. *Gen. Comp. Endocrinol.* **164**, 142–150 (2009).
10. Becker, J. B., Breedlove, S. M., Crews, D. & McCarthy, M. M. *Behavioral Endocrinology* 2nd edn (MIT Press, 2002).
11. Bonier, F., Martin, P. R., Moore, I. T. & Wingfield, J. C. Do baseline glucocorticoids predict fitness? *Trends Ecol. Evol.* **24**, 634–642 (2009).
12. Spirov, A. & Holloway, D. Using evolutionary computations to understand the design and evolution of gene and cell regulatory networks. *Methods* **62**, 39–55 (2013).

Author Contributions J.v.G. and F.J.W. both contributed to all aspects of this work.

Competing Financial Interests Declared none.

doi:10.1038/nature25495

Contesting the evidence for non-adaptive plasticity

ARISING FROM C. K. Ghalambor *et al.* *Nature* **525**, 372–375 (2015); doi:10.1038/nature15256

The effect of phenotypic plasticity on the evolution of traits is a highly controversial subject. Ghalambor *et al.*¹ added a new spin to this debate as they contrasted gene expression patterns of guppy populations evolved at two different predation levels. After four generations in a new environment, gene expression plasticity evolved with two interesting key features: (1) in environments without predation, gene expression in the evolved populations changed in a direction that decreases ancestral plasticity; and (2) the ancestral level of plasticity and the magnitude of change in plasticity are negatively correlated. However, this pattern could be an artefact of the analysis procedure rather than reflecting selection, and simple computer simulations that assume no divergence in gene expression between populations and only stochastic variation are able to replicate the pattern. There is a Reply to this Comment by Ghalambor, C. K. *et al.* *Nature* **555**, <http://dx.doi.org/10.1038/nature25497> (2018).

The analysis of genome scale data in evolutionary experiments is very challenging because it is important to distinguish genetic drift from directional forces (selection). Ghalambor *et al.*¹ approach this problem in two steps. First, they identify genes that are differentially expressed between ancestral and evolved populations in the new predator-free environment. Second, they condition on the average effect of the two evolved populations to go in the same direction as for one additional population that naturally colonized a predator-free environment. This step is included to distinguish between genetic drift and selection, because only selection is directional. Although we agree with the rationale of the testing procedure, we caution that the implementation could generate a signal of ‘concordantly differentially expressed (CDE)’ genes, even in the absence of selection.

Assuming no genetically based differences in gene expression and random differences in gene expression following a normal distribution, we simulated 37,493 transcripts and applied the testing procedure of Ghalambor *et al.*¹ We show that CDE transcripts can be detected even in the absence of selection. As in ref. 1, for CDE transcripts we also detect a negative correlation between ancestral plasticity and divergence in gene expression (Fig. 1a).

This negative correlation is the result of a statistical phenomenon known as the regression towards the mean (RTM), which was first described by Sir Francis Galton²: he observed that tall parents have, on average, children that are smaller than them (and vice versa). In this example, the parents at the extreme end of a distribution have children whose heights are closer to the mean of the distribution: that is, they regress towards the mean.

In our simulations, all transcripts have the same distribution in both environments. CDE transcripts have by chance diverged from this mean and thus were kept for further analysis. Yet, when contrasted to their expression in the second environment, they will probably regress towards the mean. Here, RTM creates the negative correlation as illustrated in Fig. 2: because three independent means are required for one group, but only one for the other, we expect that the first group will have values closer to the mean than the latter one, creating an asymmetry. Because random draws of the ancestral population under predation conditions are expected to regress towards the mean, the initial conditioning of the CDE transcripts causes the observed negative correlation (Fig. 2). Using the CDE transcripts, we also reproduce the

negative correlation between the magnitude of plasticity in the source population and the evolved mean change in plasticity (Fig. 1b). As in Fig. 1a, we think that this pattern can be explained by the conditioning in the analysis. The more extreme the expression of the ancestral population at low predator conditions is, the less deviation from the mean is needed for the evolved populations to be significant in the first step of the analysis. Hence, drawing a random sample (for the predator environment) will result in small differences between the expression values for the evolved populations.

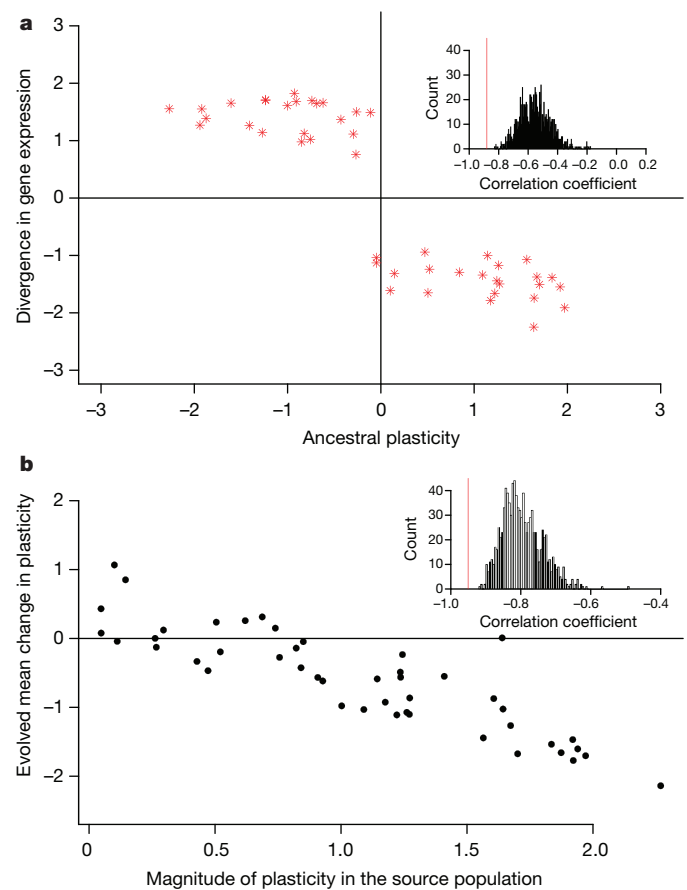


Figure 1 | Negative correlations between ancestral plasticity and divergence in gene expression, and between the magnitude of plasticity in the source population and the evolved mean change in plasticity. a, b, Negative correlation of the ancestral plasticity with the divergence in gene expression (a) and negative correlation of the mean change in plasticity in the introduced population with magnitude of plasticity in the source population (b) for CDE transcripts assuming no genetic divergence among all populations. Both insets show the distribution of Spearman's rank correlation for the CDE transcripts from 1,000 random permutations of the CDE transcripts. The correlation in the originally simulated (that is, non-permuted) dataset (red vertical line) is always more negative than in the permuted dataset.

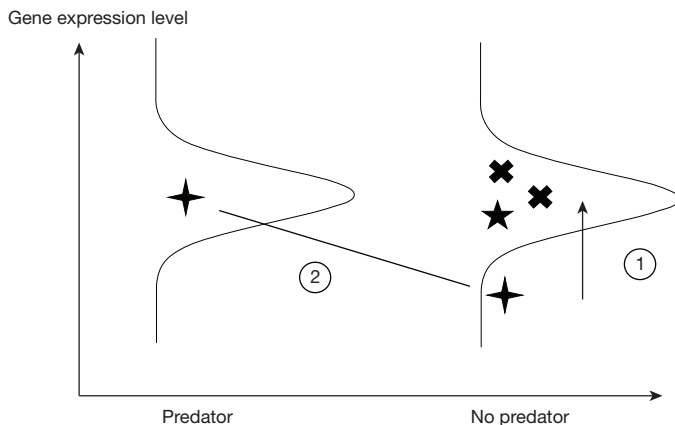


Figure 2 | Schematic representation of the negative correlation between the changes in gene expression observed in the predator-free environment and the ancestral plasticity when all values are randomly sampled from a normal distribution. The analysis of Ghalambor *et al.*¹ identifies genes expressed at one end of the distribution in the ancestral population, and closer to the mean for the three evolved ones (1). As the values in the predator environment are independent, they are also grouped around the mean and correlate negatively (2) with the evolved response: they regress towards the mean.

We showed that the key results of Ghalambor *et al.*¹ could be obtained without any genetically based difference in mean expression between the populations. Nevertheless, the results of our simulations differ in one aspect from the original study—the number of CDE transcripts in the simulated data are identical to the ones in the permuted data (all are randomly sampled). Reasoning that small differences due to sampling or even directional selection could explain this discrepancy, we performed additional simulations assuming small random changes in mean expression between populations. We confirmed that the number of CDE transcripts is higher than in the permuted dataset, exhibiting the same negative correlations without selection on plasticity (see Supplementary Information). The effect of RTM decreases with increasing between-population divergence relative to within-population variation³. For the data of Ghalambor *et al.*¹, the effect of RTM is difficult to quantify because each transcript has specific mean expression values and within and between population variances. Hence, we contrast the results of computer simulations using different parameter combinations to the experimental data (Supplementary Fig. 4).

The experimental data from Ghalambor *et al.*¹ were most similar to computer simulations in which the between-population divergence estimator was lower than the within-population variation; that is, the simulation data that show the RTM effect.

In any case, our simulations were not designed to match the experimental details of Ghalambor *et al.*¹ but instead to highlight that even without evolution of gene expression plasticity, we are able to recapitulate the key results of their study. Thus, we do not claim that the guppies have not evolved their plasticity of gene expression, but we feel that the authors would need to provide additional analyses to support their conclusions.

Methods

We randomly sampled gene expression values from a normal distribution for 37,943 transcripts in 32 samples. For each population comparison, we computed Student's *t*-test on 250 randomly permuted datasets and identified transcripts for which the simulated data were located in the 5% tails of the *t*-statistic distribution. We then calculated the mean differences in simulated expression between treatments and populations to determine the correlations between ancestral plasticity and divergence in gene expression (Fig. 1a) and magnitude of plasticity in the source population and evolved mean change in plasticity within introduced populations (Fig. 1a).

Code availability. The R code for the simulations are provided in the Supplementary Information.

François Mallard¹, Ana Marija Jakšić^{1,2} & Christian Schlötterer¹

¹Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria.

email: christian.schloetterer@vetmeduni.ac.at

²Vienna Graduate School of Population Genetics, Vienna, Austria.

Received 30 September 2015; accepted 29 January 2018.

1. Ghalambor, C. K. *et al.* Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**, 372–375 (2015).
2. Galton, F. Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. G. B. Irel.* **15**, 246–263 (1886).
3. Barnett, A. G., van der Pols, J. C. & Dobson, A. J. Regression to the mean: what it is and how to deal with it. *Int. J. Epidemiol.* **34**, 215–220 (2005).

Supplementary Information accompanies this Comment.

Author Contributions F.M. performed simulations and analysed the data. F.M., A.M.J. and C.S. conceived the study and wrote the manuscript.

Competing Interests Declared none.

doi:10.1038/nature25496

Ghalambor *et al.* reply

REPLYING TO J. van Gestel & F. J. Weissing *Nature* **555**, <http://dx.doi.org/10.1038/nature25495> (2018); F. Mallard, A. M. Jakšić & C. Schlötterer *Nature* **555**, <http://dx.doi.org/10.1038/nature25496> (2018)

The concerns raised in the accompanying Comments by Mallard *et al.*¹ and van Gestel and Weissing² that natural selection could not generate the transcriptomic results reported in our Letter³ overlook the larger context of previous work documenting rapid parallel evolutionary changes in guppies^{4–7}. Here we also show compelling evidence that their alternative interpretations simply do not match our published datasets.

Mallard *et al.*¹ argue that the negative correlation between plasticity and evolution reflects neutral processes rather than selection, but they only report simulations using a subset of criteria our conclusions were based on; when all criteria are applied, their simulations support our conclusions. We based our conclusions on four criteria with analyses that appropriately accounted for stochastic variation and non-independence of our data: (1) we found more concordantly differentially expressed (CDE) transcripts than in permuted datasets; (2) we did not find more genes that diverged in opposite directions than in permuted datasets; (3) the association between plasticity and divergence was more negative than in permuted datasets; and (4) the direction of association between plasticity and divergence was more extreme than in permuted datasets. Had our data not met both of the first two criteria, we would not have concluded that our CDE genes were enriched for genes evolving under selection and would not have proceeded to assess the relationship between ancestral plasticity and divergence. The results presented in figures 1 and 2 of Mallard *et al.*¹ do not meet criterion (1) and thus provide no evidence that weakens our conclusions. The authors acknowledge this potential flaw and present simulations in their supplementary methods that meet criterion (1), yet they omit criterion (2) in those analyses. In fact, only 26 of their 400 parameter sets replicate all four criteria, and only a fraction of these yield datasets with distributions that might reasonably match our results (Fig. 1 and ref. 8). Therefore, the simulations of Mallard *et al.*¹ are not only

consistent with our interpretation, but also other studies concluding that rapid evolution of genes exhibiting non-adaptive plasticity is more likely due to selection and very unlikely to arise by chance^{3,9}.

We agree with van Gestel and Weissing² that individual transcripts are not independent, and recognize the robustness of gene expression networks to produce similar phenotypes via different mechanisms¹⁰. We explicitly incorporated non-independence among transcripts in designing our permutations to preserve entire transcriptional profiles. The model proposed by van Gestel and Weissing² positing a single regulatory change underlying patterns in our set of CDE transcripts simply does not match our data, as such a model would generate a strongly correlated set of CDE transcripts. Correlational analyses show that few of our CDE genes are strongly correlated, and that numerous clusters of transcripts independently evolved CDE expression patterns and negative associations between plasticity and divergence (see ref. 8) to compare correlations in experimental data and simulated datasets with a single regulatory change). We thus refute the claim that variation in a single modulator can account for the rapid evolution of our CDE transcripts, just as it is unlikely to explain the rapid parallel evolution of other complex phenotypes in guppies.

Cameron K. Ghalambor^{1,2*}, Kim L. Hoke^{1,2*}, Emily W. Ruell¹, Eva K. Fischer¹, David N. Reznick³ & Kimberly A. Hughes⁴

¹Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA.

email: cameron1@colostate.edu

²Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado 80523, USA.

³Department of Biology, University of California, Riverside, California 92521, USA.

⁴Department of Biological Science, Florida State University, Tallahassee, Florida 32306-4295, USA.

*These authors contributed equally to this work.

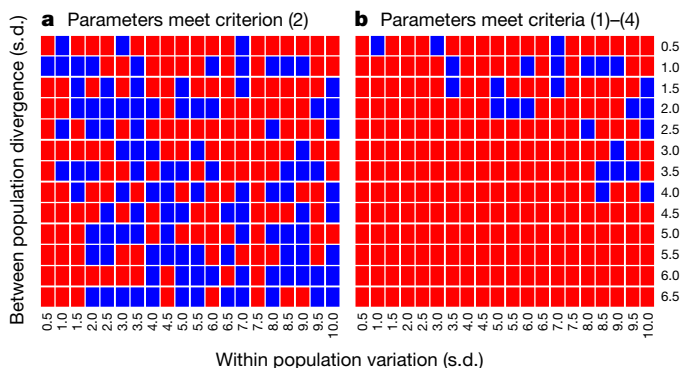


Figure 1 | A small fraction of simulations performed by Mallard *et al.*¹ meet all four criteria that support our conclusion that CDE genes diverged under selection, and that they diverged in the opposite direction as ancestral plasticity. Blue boxes indicate parameter sets that met the criteria, and red boxes indicate parameter sets that did not meet the criteria. **a, Only a subset of parameters meet criterion (2), positing that differentially expressed genes that diverge in opposite directions should not be overrepresented compared to the permuted datasets. **b**, All four criteria merged together shows that only 26 parameter sets recreate the results found in Ghalambor *et al.*³ This contrasts with the much higher fraction claimed by Mallard *et al.*¹, as they did not consider criterion (2) in their analyses.**

- Mallard, F., Jakšić, A. M. & Schlötterer, C. Contesting the evidence for non-adaptive plasticity. *Nature* **555**, <http://dx.doi.org/10.1038/nature25496> (2018).
- van Gestel, J. & Weissing, F. J. Is plasticity caused by single genes? *Nature* **555**, <http://dx.doi.org/10.1038/nature25495> (2018).
- Ghalambor, C. K. *et al.* Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**, 372–375 (2015).
- Reznick, D. A., Bryga, H. & Endler, J. A. Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).
- Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**, 1934–1937 (1997).
- Handelsman, C. A. *et al.* Predator-induced phenotypic plasticity in metabolism and rate of growth: rapid adaptation to a novel environment. *Integr. Comp. Biol.* **53**, 975–988 (2013).
- Gordon, S. P. *et al.* Selection analysis on the rapid evolution of a secondary sexual trait. *Proc. R. Soc. B* **282**, 20151244 (2015).
- Hoke, K. L. *et al.* Untangling the role of selection and drift in population divergence via transcriptional network simulations. Preprint at <https://doi.org/10.1101/277830> (2018).
- Ho, W.-C., & J. Zhang. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. *Nat. Comm.* **9**, 350 (2018).
- Fischer, E. K., Ghalambor, C. K. & Hoke, K. L. Can a network approach resolve how adaptive and non-adaptive plasticity impacts evolutionary trajectories? *Integr. Comp. Biol.* **56**, 877–888 (2016).

doi:10.1038/nature25497

Oceans on Mars formed early

The geometry of putative ancient shorelines on Mars suggests that these features were deformed by the growth of a massive volcanic region — a finding that has implications for the climate, geology and hydrology of early Mars. [SEE LETTER P.643](#)

MARIA T. ZUBER

Arguably the most compelling evidence that Mars once had oceans is the presence of potential shorelines in the planet's northern hemisphere¹. However, these features do not have a constant elevation, which would argue against an oceanic origin². On page 643, Citron *et al.*³ show that the deviations can be explained if the shorelines formed before and during the growth of Tharsis — a volcanic region that spans nearly one-quarter of Mars's surface area. The results imply that the formation of the planet's oceans occurred earlier than previously thought, and might be linked to the volcanism of Tharsis and the formation of valley networks⁴.

The history of liquid water on Mars is of great interest because water would have strongly influenced the planet's climate and its potential for habitability. Evidence for a watery past in Mars's geological record is abundant, but many of the data are indicative of catastrophic outflows or intermittent releases that neither require a temperate surface environment, nor are thought to have persisted long enough for the development of microbial life⁵.

Researchers have long debated the possibility that a liquid-water ocean or oceans once filled the topographic depression in Mars's northern hemisphere⁶. Comparisons of the roughness⁷ and intrinsic electrical properties (dielectric properties)⁸ of the proposed Martian sea floor with those of similarly sediment-laden surfaces on Earth, as well as the concentrations of certain isotopes on Mars⁹, support the idea of past oceans. But perhaps the most convincing evidence comes from geomorphological observations of potential ancient shorelines. In particular, geological mapping of spacecraft images has revealed putative shorelines that can be tracked for thousands of kilometres on Mars's northern plains¹.

The identification of these shorelines is not without caveats. For instance, the geomorphological observations are open to interpretation². Furthermore, shorelines, like beaches on Earth, should trace surfaces of constant gravitational potential (such as sea levels). However, Mars's putative shorelines show changes in elevation of up to several kilometres, indicating

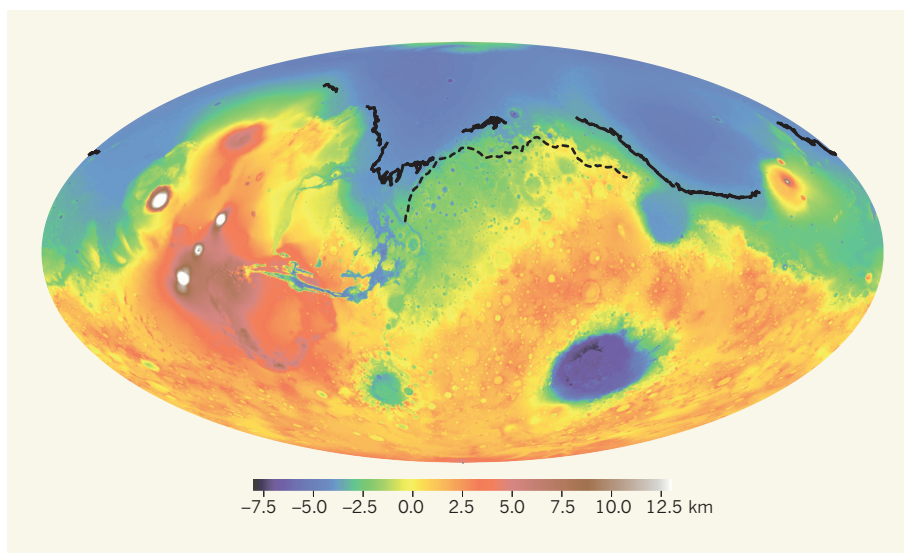


Figure 1 | Mars's topography and putative ancient shorelines. The surface of Mars contains a massive region called Tharsis (red), which was formed by volcanic activity, and lowland areas in the northern hemisphere (blue) that have been proposed to be the location of ancient oceans⁵. Geological mapping¹ has revealed two extensive potential past shorelines: Arabia (dashed line) and Deuteronilus (solid lines). Citron *et al.*³ report that the geometry of these shorelines can be explained if they were deformed in response to the growth of Tharsis, rather than after Tharsis's formation as previously assumed. The colour scale denotes elevation in kilometres. (Adapted from Extended Data Fig. 2 of the paper³.)

that if these features indeed formed at ocean coasts, they subsequently underwent substantial deformation. If the event that produced such deformation could be identified, it would constrain the relative timing of the formation of the shorelines, and therefore of the oceans.

Citron and colleagues use shoreline modelling to show that the observed deformation can be explained by the growth of Tharsis, most of which occurred some 3.6 billion or more years ago, during an early stage in Mars's history called the Noachian period. The authors considered two extensive shorelines: Arabia and Deuteronilus (Fig. 1). Their analysis indicates that the Arabia shoreline formed before or in the early stages of Tharsis's growth, whereas the Deuteronilus shoreline arose during the late stages. These findings are in contrast to general thinking that the oceans, if they existed, formed after Tharsis's emplacement. By correcting for deformation events, Citron *et al.* place lower limits on the ocean volumes associated with the Arabia and Deuteronilus shorelines, of about 11 and 3 times that of

Earth's Mediterranean Sea, respectively.

A previous study¹⁰ showed that the elevation patterns of the Arabia and Deuteronilus shorelines could be explained by the deformation induced by a reorientation of Mars's spin axis — a phenomenon known as true polar wander¹¹. Rotational dynamics dictates that a planet spins stably about its axis of minimum inertia. Consequently, the load associated with an ocean of the calculated volume could have resulted in true polar wander if Tharsis had formed far from Mars's equator. However, evidence now indicates that Tharsis probably originated near the equator^{12,13}, in which case, ocean loading would not have been great enough to destabilize the planet's spin axis. If Tharsis had formed close to, but not precisely at, the equator, loading would have induced a limited amount of true polar wander, but not enough to explain the deformation of the proposed shorelines.

Citron and colleagues' scenario is therefore more consistent with our current understanding of Tharsis's emplacement. The

authors' revised sequence of events also has some appeal, in that it ties ocean formation to the release of water-rich gas from massive, extended volcanic events. And, in the case of the ocean associated with the Arabia shoreline, it links ocean formation to the emergence of valley networks⁴ — abundant features on early Mars that carried water to the northern lowlands in which the oceans were situated.

Not surprisingly, the interplay of water and geodynamics over Mars's history is more complicated than presented in Citron and colleagues' analysis. During the period of interest, additional deformational processes occurred at local to global length scales that were not accounted for. Such processes are associated with, for example, mantle activity, impacts of asteroids or comets on to Mars's surface, glaciers, erosion and individual volcano growth, and could also have contributed to the deformation of the shorelines.

More-detailed mapping and modelling could clarify the contributions to the observed elevation changes and refine the sequence and relative timing of events. In addition, other, much shorter and more fragmented potential shorelines have been proposed^{14,15} that suggest the existence of seas and lakes, and that merit further study. Finally, the timing and role of volcanism, as it relates to the state and evolution of the atmosphere and to the persistence of water on the planetary surface, is far from understood. But, fortunately, the extensive database of remote and *in situ* observations of Mars continues to accumulate, revealing the intriguing story of Mars's water-rich past. ■

Maria T. Zuber is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: zuber@mit.edu

- Parker, T. J., Gorsline, D. S., Saunders, R. S., Pieri, D. C. & Schneeberger, D. M. *J. Geophys. Res. Planets* **98**, 11061–11078 (1993).
- Malin, M. C. & Edgett, K. S. *Geophys. Res. Lett.* **26**, 3049–3052 (1999).
- Citron, R. I., Manga, M. & Hemingway, D. J. *Nature* **555**, 643–646 (2018).
- Bouley, S. *et al. Nature* **531**, 344–347 (2016).
- Carr, M. H. *Water on Mars* 197 (Oxford Univ. Press, 1996).
- Baker, V. R. *et al. Nature* **352**, 589–594 (1991).
- Head, J. W. III *et al. Science* **286**, 2134–2137 (1999).
- Mouginot, J., Pommerol, A., Beck, P., Kofman, W. & Clifford, S. M. *Geophys. Res. Lett.* **39**, L02202 (2012).
- Villanueva, G. L. *et al. Science* **348**, 218–221 (2015).
- Perron, J. T., Mitrova, J. X., Manga, M., Matsuyama, I. & Richards, M. A. *Nature* **447**, 840–843 (2007).
- Goldreich, P. & Toomre, A. *J. Geophys. Res.* **74**, 2555–2567 (1969).
- Roberts, J. H. & Zhong, S. *Icarus* **190**, 24–31 (2007).
- Matsuyama, I. & Manga, M. *J. Geophys. Res. Planets* **115**, E12020 (2010).
- Clifford, S. M. & Parker, T. J. *Icarus* **154**, 40–79 (2001).
- Carr, M. H. & Head, J. W. III *J. Geophys. Res. Planets* **108**, 5042 (2003).

SENSORY SYSTEMS

A trio of ion channels takes the heat

Of the various temperature-sensitive ion channels identified previously, three have now been found to act in concert to detect painful heat and initiate protective reflexes. [SEE LETTER P.662](#)

ROSE Z. HILL & DIANA M. BAUTISTA

There's nothing cosier than sitting by the fire on a cold night, and mammals have an amazing ability to find the perfect spot in which to warm themselves without being burnt. This ability is mediated by specialized heat-sensitive neurons of the somatosensory system that innervate the skin. These neurons are fine-tuned to detect the temperature at which pleasurable warmth turns into painful (noxious) heat, and are responsible for initiating protective reflexes and triggering avoidance behaviours. The search for temperature sensors in somatosensory neurons has uncovered several ion channels that are part of the transient receptor potential (TRP) family^{1–4}, but the TRP channel or channels that are responsible for heat-evoked pain have remained unclear. Vandewauw *et al.*⁵ report on page 662 that three TRP channels work together to detect noxious heat in mice.

The first temperature-sensitive TRP channel, TRPV1, was discovered more than 20 years ago¹. TRPV1 has a crucial role in the development of hypersensitivity to heat after injury or inflammation, but experiments have revealed that mice that lack TRPV1 display only a partial defect in the ability to sense and respond to noxious heat⁶. Another TRP

channel in somatosensory neurons, TRPM3, is also activated in response to painful temperatures; however, heat-evoked pain-avoidance behaviours are only partially attenuated in mice that lack this channel³, leaving the mystery of how painful heat is sensed unsolved.

Vandewauw *et al.* report that double-mutant mice, which lack both the TRPV1 and TRPM3 channels, show only mild defects in heat-evoked pain-avoidance behaviours, similar to mice that lack just one of those channels. By contrast, they find that mice that lack a third channel, TRPA1, as well as TRPV1 and TRPM3, do not display any protective avoidance behaviours when exposed to noxious heat (Fig. 1). The three channels therefore act in concert to mediate behavioural responses to noxious heat.

What led Vandewauw and colleagues to implicate TRPA1 in heat sensing? The authors observed that a subset of the heat-responsive somatosensory neurons that normally express TRPV1 and TRPM3 also express TRPA1, which has previously been shown to mediate inflammatory pain⁷ and itching⁸ in mice. TRPA1 is known to contribute to the heat-sensing abilities of non-mammalian organisms, including rattlesnakes⁹ and fruit flies¹⁰, but mice that lack TRPA1 exhibit normal

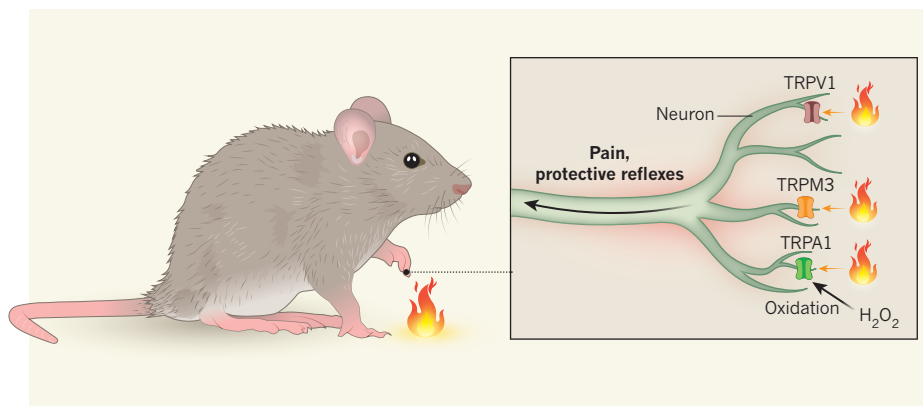


Figure 1 | Three heat-sensitive TRP channels. Vandewauw *et al.*⁵ demonstrate in mice that potentially damaging temperatures excite pain-sensing neurons of the peripheral nervous system through the activity of three ion channels from the TRP family — TRPV1, TRPM3 and TRPA1. Activation of TRPA1 (by, for example, hydrogen peroxide, H₂O₂) is likely to require oxidation of certain amino-acid residues, as well as the presence of heat. Activation of the three channels causes neuronal firing and the transmission of signals to the central nervous system, leading to pain-associated behaviours and protective reflexes.

avoidance of noxious heat^{7,11}. Therefore, the protein has not been implicated in heat sensing in mammals — until now.

Activation of this channel by heat in rattlesnakes or fruit flies involves specialized sequences of amino acids known as ankyrin repeats at the amino terminus¹². These repetitive stretches are different in mammalian TRPA1, making the channel insensitive to heat under normal conditions¹². By contrast, activation of TRPA1 in mice by inflammatory molecules is dependent on the oxidation state of three cysteine amino-acid residues located adjacent to the ankyrin repeats¹³. Vandewauw and colleagues therefore proposed that oxidation of these residues might be sufficient to activate the heat-sensing capabilities of mouse TRPA1.

To trigger oxidation of TRPA1 *in vitro*, the authors applied hydrogen peroxide to sensory neurons that lacked TRPV1 and TRPM3 but expressed TRPA1. They found that hydrogen peroxide treatment did sensitize TRPA1 in such a way that heat could now excite the neurons. This suggests that mouse TRPA1 must be oxidized to respond to heat. However, the authors do not provide a physiological mechanism for how TRPA1 might be oxidized *in vivo* to participate in heat sensing. This will no doubt be an area for future investigation.

Vandewauw and colleagues also performed electrophysiological experiments, which revealed that the three TRP channels they studied are co-expressed in more than one class of somatosensory nerve fibre — including at least two types of pain-sensing neuron that encode distinct sensations, such as dull pain or sharp pain. The expression of these channels in different neuronal subtypes might give rise to complex circuitry and crosstalk that aids the rapid avoidance of noxious heat. In addition, it could add a further layer of redundancy to an important protective response.

Finally, the group showed that, although the triple-mutant mice do not exhibit protective heat-avoidance reflexes, they do have the same preference as control mice for innocuous warm temperatures (30 °C) over noxious high temperatures (45 °C) when presented with a choice. This finding suggests that the sensing of painful heat might not have a substantial role in determining a preference for pleasurable temperatures over painful ones, warranting further investigation of warmth sensing — an area of active debate. A previous study has shown that whereas heat-responsive spinal neurons, which receive input from heat-sensitive somatosensory neurons, respond to absolute temperatures, cold-responsive spinal neurons, which receive input from cold-sensitive somatosensory neurons, respond to relative changes in temperature¹⁴. Together, these data imply that the main role of peripheral heat-sensing neurons, such as those described by Vandewauw *et al.*, might be

to mediate avoidance of high temperatures, rather than to set an animal's preferred temperature range.

Circuit-tracing and *in vivo* imaging will no doubt reveal the contributions of different subsets of peripheral neurons to the various representations of cold, cool, warm and hot temperatures in the central nervous system. Decoding these circuits will help to unravel the nuances of sensory perception and to elucidate the basis for diverse temperature preference between species or individual organisms. ■

Rose Z. Hill and Diana M. Bautista are in the Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. D.M.B. is also at the Helen Wills Neuroscience Institute, University of California, Berkeley.

e-mails: rzhill@berkeley.edu; dbautista@berkeley.edu

1. Caterina, M. J. *et al.* *Nature* **389**, 816–824 (1997).
2. Patapoutian, A., Peier, A. M., Story, G. M. & Viswanath, V. *Nature Rev. Neurosci.* **4**, 529–539 (2003).
3. Vriens, J. *et al.* *Neuron* **70**, 482–494 (2011).
4. Tan, C. H. & McNaughton, P. A. *Nature* **536**, 460–463 (2016).
5. Vandewauw, I. *et al.* *Nature* **555**, 662–666 (2018).
6. Caterina, M. J. *et al.* *Science* **288**, 306–313 (2000).
7. Bautista, D. M. *et al.* *Cell* **124**, 1269–1282 (2006).
8. Wilson, S. R. *et al.* *Nature Neurosci.* **14**, 595–602 (2011).
9. Gracheva, E. O. *et al.* *Nature* **464**, 1006–1011 (2010).
10. Neely, G. G. *et al.* *PLoS ONE* **6**, e24343 (2011).
11. Kwan, K. Y. *et al.* *Neuron* **50**, 277–289 (2006).
12. Cordero-Morales, J. F., Gracheva, E. O. & Julius, D. *Proc. Natl Acad. Sci. USA* **108**, E1184–E1191 (2011).
13. Macpherson, L. J. *et al.* *Nature* **445**, 541–545 (2007).
14. Ran, C., Hoon, M. A. & Chen, X. *Nature Neurosci.* **19**, 1201–1209 (2016).

This article was published online on 14 March 2018.

CHEMISTRY

AI designs organic syntheses

Software that devises effective schemes for synthetic chemistry has depended on the input of rules from researchers. A system is now reported in which an artificial-intelligence program learns the rules for itself. [SEE ARTICLE P.604](#)

DEREK LOWE

Synthetic organic chemistry is the science of building desired chemical structures from simpler parts. The knowledge and experience of researchers has always been the key to combining chemical reactions into successful synthetic schemes. But on page 604, Segler *et al.*¹ report that an artificial-intelligence program can design routes for synthesizing compounds that — at least, on paper — seem just as good as those produced by humans.

Organic chemists often work by thinking backwards as much as they do forwards when designing a synthetic route. The concept of retrosynthesis², introduced by E. J. Corey in the 1960s, and for which he was awarded the Nobel Prize in Chemistry in 1990, codified the way in which many chemists think (Fig. 1). When looking at a target molecule, they ask: “What could this have been made from? Which bonds could have been formed, and which atoms or chemical groups could have been added or transformed?” Then, the process starts again, as researchers try to determine the reactions that could have led to the precursor molecule. The aim is to work back to easily available starting compounds, while balancing the

factors that make a good synthesis — including the number of steps involved, the probable product yields of those steps, and the ease of use of the chemistry involved. Organic chemists deal constantly with such questions, for example when making compounds for testing in drug-discovery programmes.

Since the birth of synthetic organic chemistry in the mid-nineteenth century, a huge number of synthetic organic reactions have been reported across a literature that gets larger every hour. Before the 1980s, many chemists kept collections of handwritten, cross-referenced index cards containing useful reactions from the literature, to guide the design of synthetic pathways. These aide-memoires moved naturally on to digital databases as computer technology became widespread.

These days, chemists review the various methods for turning chemical group X into chemical group Y by drawing the molecular structures of interest using a computer program and then performing an online search for relevant reactions. This almost invariably produces a long list, from which researchers must select the most appropriate reaction for their needs, according to their knowledge and experience. Stringing such reactions into a useful

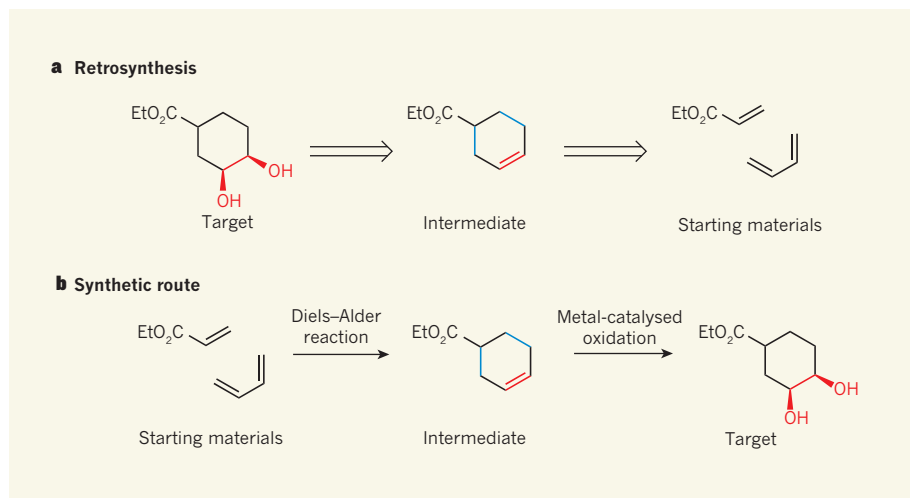


Figure 1 | Retrosynthetic analysis. Chemists use a strategy known as retrosynthesis to design schemes for constructing organic molecules. **a**, The first step is to work out a theoretical pathway by which a target molecule is sequentially converted into intermediates that can be easily prepared or purchased. In this example, the red bonds and atoms in the target can be derived from the red double bond in the intermediate molecule. The red double bond and the blue bonds in the intermediate can, in turn, be formed in a single step from two commercially available starting materials (Et, ethyl group). **b**, The second step is to choose reactions that will convert the starting materials into the target molecule: here, the Diels–Alder reaction forms the intermediate, which undergoes metal-catalysed oxidation to form the target. Segler *et al.*¹ report that an artificial-intelligence program has devised plausible synthetic routes for molecules by inferring design rules on its own.

synthesis has been thought of as a problem that only humans can solve.

But does it have to be? Could a sufficiently large and well-curated database of chemical transformations be used as the basis for a program that not only finds reactions, but also arranges them into plausible synthetic plans? Such programs have been sought since Corey's work in the 1960s, but (until recently) with little practical success.

Two fundamental problems have frustrated the dream. First, computing hardware simply could not tackle the scale of the challenge. Second, the chemical literature is hard to define in terms that a software program can understand: given reactions would work for the type of compound for which they were claimed to work (most of the time), but only under certain conditions. For example, group X would turn into group Y, unless group Z was present elsewhere in the reactant molecules. When group Z was present, the reaction might still work if group Q was nearby in the same molecule — but only, for instance, when the pH was lower than a certain value, or when the temperature was high enough, or when there was no water present.

There are various ways to overcome this second problem. One is to provide the program with an exhaustively human-curated list of the reactions that can enable a desired chemical transformation, and which takes into account all the limitations and conditions. The program can then combine such reactions into synthetic routes in ways that are broadly similar to those used to evaluate combinations of chess moves. This approach is starting to yield

results (see ref. 3, for example), and several competing commercial software products are available.

Segler *et al.* have investigated another method: instead of getting researchers to load their expertise into a machine, is it possible to design a program that learns by itself what researchers know? This concept has already produced startling results, with programs

“The AI-generated synthetic routes were evaluated by trained organic chemists in a blind test for plausibility.”

that can learn to play games such as Go on their own⁴, rather than being trained using lists of human strategies. The authors devised a computational process that starts by automatically extracting chemical transformations from a large commercial database, being careful to include only reactions that have been reported several times. Their system accepts these well-precedented reactions as ‘allowed moves’ in organic synthesis. When the system is asked to devise a synthetic route to a target molecule, it works backwards from the target as would a human, picking out the most promising precursor molecules according to the design rules that it has learnt, and then seeing how feasible it is to synthesize those. The authors combined three artificial neural networks with a random Monte Carlo tree search — a type of search algorithm used by computers in certain decision-making processes — to narrow down the most

promising synthetic routes, without getting stuck too quickly on a particular path.

Importantly, the routes that emerged were evaluated not only by the program's scoring system, but also by trained organic chemists in a blind test for plausibility. When the chemists were asked to assess machine-generated synthetic pathways for target molecules alongside routes reported in the literature, they expressed no preference for the routes that had been shown to work by their fellow researchers. In other words, they found the chemistry suggested by the program to be as reasonable as the syntheses proposed by researchers.

This does not necessarily mean that all machine-suggested routes will work in the laboratory — but, as organic chemists know to their sorrow, many routes designed by humans fail there, too. Further development of the program could include such ‘reductions to practice’, to determine whether the machine-proposed routes are better (or, at least, no worse) than those devised by people. A study³ this year that assessed a more conventional, hand-curated retrosynthesis program is notable for its inclusion of such a laboratory test. Achieving routes that are ‘no worse’ than those of researchers is a clear victory for Segler and colleagues' program, which arrives at pathways in considerably less time, and with much greater coverage of the literature, than a person could manage.

If such programs fulfil their promise, and there is little reason to think that they won't, synthetic chemists will find that a mainstay of their work starts to disappear. Technological innovations have had similar effects in the past, but usually by automating physical ‘grunt work’ that is missed by no one. Disconcertingly, developments in artificial intelligence encroach on the thinking part of the job. There will always be complex, unusual and unprecedented structures that such software cannot handle, but the task of solving more-routine synthetic questions will be taken out of the hands of researchers.

The idea that intellectual tasks can be categorized as automatable grunt work will probably be insulting to many chemists, and will certainly feel like a threat. But the use of artificial intelligence will actually free up time in which they can think more about the higher-level questions of which molecules should be made and why, rather than focusing on the details of how to make molecules. Not all researchers will welcome this shift. But it seems to be coming, regardless. ■

Derek Lowe is a science writer based in Boston, Massachusetts.
e-mail: derek.lowe@gmail.com

1. Segler, M. H. S., Preuss, M. & Waller, M. P. *Nature* **555**, 604–610 (2018).
2. Corey, E. J. *Angew. Chem. Int. Edn* **30**, 455–465 (1991).
3. Klucznik, T. *et al. Chem* **4**, 522–532 (2018).
4. Silver, D. *et al. Nature* **550**, 354–359 (2017).

MATERIALS SCIENCE

Wing origami

Origami-based design techniques have applications ranging from nanoscale devices to architectural structures. But conventional techniques have a limited range of folding patterns. Writing in *Science*, Faber *et al.* report an approach that overcomes this restriction (J. A. Faber *et al. Science* **359**, 1386–1391; 2018).

The authors were inspired by the amazing folding abilities of earwig wings (pictured). They discovered that protein-rich joints in the wings behave like springs, and mimicked this design in synthetic structures using mechanical springs. Such structures show folding patterns and functionalities that were impossible using previously available methods. [Ryan Wilkinson](#)



PANTHER MEDIA GMBH/ALAMY

IMMUNOLOGY

Killer T cells show their kinder side

The immune system protects the body by responding to invading organisms. But how is an attack on useful resident microbes prevented? A pathway has now been identified that allows immune cells to sense and respond to beneficial bacteria.

PAUL KLENERMAN & GRAHAM OGG

The immune system performs a difficult balancing act. It must respond rapidly to dangerous microorganisms that threaten the host, yet also coexist peacefully with the large array of microorganisms that colonize barrier surfaces such as those of the skin and gut. Writing in *Cell*, Linehan *et al.*¹ examine the interactions between immune cells and a bacterium that normally resides on the skin, and identify a signalling pathway that governs the immune response to such microbes.

Most immunological research has focused on how immune cells recognize and eliminate viral and bacterial pathogens, and the insights obtained have led to the development of many immune-based therapies. However, attention has also turned to microbial communities at barrier surfaces and the long-term influence they exert on host immune responses. Accordingly, Linehan and colleagues studied the skin-dwelling bacterium *Staphylococcus epidermidis* in mice and rhesus monkeys, as well as analysing human-tissue samples from the clinic.

Staphylococcus epidermidis belongs

to a group of microorganisms termed commensals, and usually colonizes human skin without causing disease. However, in premature newborns and people who are in an immunosuppressed state or who have received artificial joints, *S. epidermidis* can cause severe disease or chronic infection². Furthermore, even though the bacterium is normally harmless, it is not ignored by the immune system. In mouse skin, the response³ to an initial encounter with certain strains of *S. epidermidis* is dominated by a type of immune cell called a cytotoxic CD8⁺ T cell, also known as a killer T cell. Linehan and colleagues asked two key questions about this response, and the answers they obtained are illuminating.

The first question was, how do T cells recognize *S. epidermidis*? A response to bacteria is thought to occur when a T cell recognizes a bacterial peptide fragment bound to the 'classical' class Ia protein component of the major histocompatibility complex (MHC), which presents such fragments to immune cells. However, using a mouse model system, Linehan and colleagues found that a 'non-classical' protein, H2-M3, which belongs to the MHC class Ib family, was instead involved in MHC presentation in the response

to *S. epidermidis*. H2-M3 can present⁴ bacterial peptides in which the amino-acid residue methionine has been modified by addition of an N-formyl group.

The other question Linehan and colleagues sought to answer was whether this non-classical immune-recognition pathway might also be coupled to 'unconventional' T-cell behaviour. The team investigated this by comparing gene-expression patterns of CD8⁺ T cells responding to *S. epidermidis* with those of other types of T-cell population in the mouse skin that did not respond to the bacterium. The authors found that the responsive CD8⁺ T cells expressed genes associated with wound healing and tissue repair (Fig. 1). This was surprising because responding CD8⁺ T cells would be expected to express genes associated with activation of the pathways involved in killing unwanted cells or in the release of pro-inflammatory signalling molecules.

Using an *in vivo* mouse model of skin-wound healing in association with *S. epidermidis* colonization, the authors demonstrated that the tissue-repair gene-expression signature in responding T cells was linked to promotion of a healing response. Thus, immune recognition of bacterial residents of the skin can induce a form of immunity that does not directly retaliate against the bacterium, but instead aids tissue repair.

The authors' finding adds to a growing recognition of the diverse functions of other unconventional T cells that are found at barrier sites and that can recognize evolutionarily conserved microbial molecules. These cells include, for example, mucosal-associated invariant T cells, which are abundant in humans and induced by commensals, and which share several features⁵ of the T-cell population studied by Linehan and colleagues — such as the production of the signalling molecule interleukin-17A.

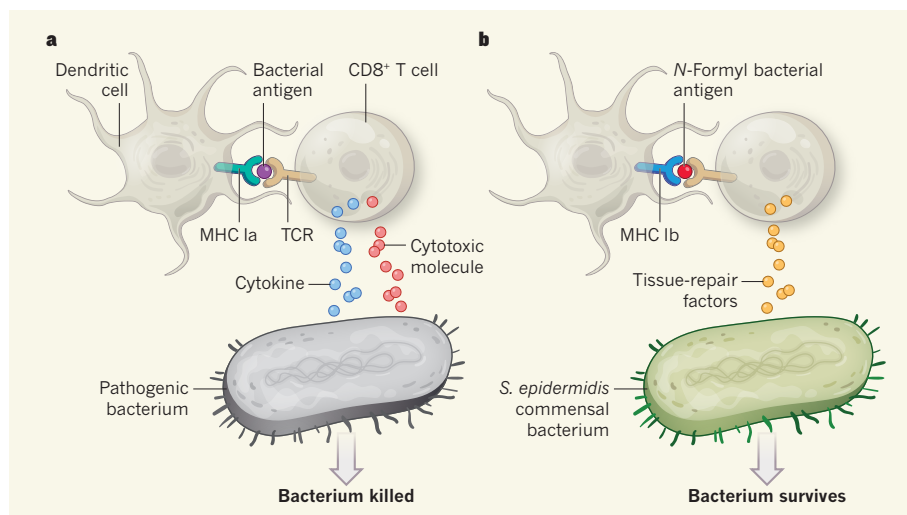


Figure 1 | Response of killer T cells to bacterial infection. **a**, An immune cell called a CD8⁺ T cell (killer T cell) can help to control infection by a pathogenic bacterium. The T-cell receptor (TCR) recognizes infection when a bacterial peptide called an antigen is presented on dendritic cells of the immune system by an MHC class Ia protein. A conventional T-cell response is then unleashed, and the release of cytokine and cytotoxic molecules triggers inflammation and bacterial destruction. **b**, Linehan *et al.*¹ report an analysis of the killer-T-cell response to *Staphylococcus epidermidis* using mouse models and human cells. This type of bacterium, called a commensal, can become a long-term, non-pathogenic resident of the body. Killer T cells recognize *S. epidermidis* when an MHC class Ib molecule presents an antigen containing an N-formyl modification of an amino-acid residue. The T cell expresses genes associated with tissue repair, and the bacterium survives this 'unconventional' response. However, such immune responses are probably not neatly divided into these categories, and conventional T-cell responses might in some cases aid tissue repair and unconventional T-cell responses might aid host defence against pathogens.

These and other T-cell populations in humans and mice, such as invariant natural killer T cells and germline-encoded mycolyl-reactive T cells, also act in response to MHC class Ib-mediated bacterial-peptide recognition^{6,7}. A human equivalent to mouse H2-M3 has not been identified so far. It is possible that, if such an equivalent is found, a human T-cell population might fulfil a similar role to that of the CD8⁺ T cells observed by Linehan and colleagues.

As the authors note, many of the day-to-day interactions between immune-system cells and

microbes are probably with the commensals in the skin, gut and airways. Therefore, an immune response that maintains the status quo, rather than driving bacterial elimination, which is usually accompanied by inflammation and tissue damage, could be beneficial. But should such a response really be considered unconventional? Although the roles of classical MHC molecules and responding immune cells were identified first, non-classical MHC and MHC-like proteins are evolutionarily older^{8,9} than MHC class Ia molecules. Linehan and

colleagues reveal a direct link between the role of MHC class Ib molecules and tissue homeostasis. Therefore, these functions might have evolved in parallel and represent the broadly 'conventional' functions of the immune system, with additional immune-system specializations evolving over time.

The authors' study hints at the potential broad, long-term effects of microbial residents on the immune system. If future studies mechanistically link individual microorganisms to specific types of immune response, it might be possible not only to understand the roles of some abundant and relatively unexplored immune-cell populations, but also to determine how resident microbes affect health and disease in a way that might lead to new treatment approaches. ■

Paul Klenerman and Graham Ogg are in the NIHR Biomedical Research Centre, Oxford OX1 3SY, UK. **P.K.** is also in the Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford. **G.O.** is also in the MRC Human Immunology Unit, University of Oxford.
e-mails: paul.klenerman@ndm.ox.ac.uk; graham.ogg@ndm.ox.ac.uk

1. Linehan, J. L. *et al.* *Cell* **172**, 784–796 (2018).
2. Sabatés Brescó, M. *et al.* *Front. Microbiol.* **8**, 1401 (2017).
3. Naik, S. *et al.* *Nature* **520**, 104–108 (2015).
4. Wang, C.-R. *et al.* *Cell* **82**, 655–664 (1995).
5. Legoux, F., Salou, M. & Lantz, O. *Annu. Rev. Cell Dev. Biol.* **33**, 511–535 (2017).
6. Van Rhijn, I. *et al.* *Nature Immunol.* **14**, 706–713 (2013).
7. Godfrey, D. I., Uldrich, A. P., McCluskey, J., Rossjohn, J. & Moody, D. B. *Nature Immunol.* **16**, 1114–1123 (2015).
8. Rodgers, J. R. & Cook, R. G. *Nature Rev. Immunol.* **5**, 459–471 (2005).
9. Goyos, A., Sowa, J., Ohta, Y. & Robert, J. J. *Immunol.* **186**, 372–381 (2011).

This article was published online on 26 March 2018.

METABOLISM

The healthy diabetic cavefish conundrum

Some Mexican cavefish have a mutation in an insulin receptor protein that affects blood-glucose regulation. The same mutation causes diabetes and health problems in humans, but the diabetic cavefish thrive. SEE LETTER P.647

SYLVIE RÉTAUX

Blind Mexican cavefish (*Astyanax mexicanus*) live in dark caves, a challenging environment that has shaped their evolution. On page 647, Riddle *et al.*¹ reveal the surprising way in which cavefish regulate their

glucose metabolism, a finding that might have implications for our understanding of human diabetes.

Blood-glucose regulation is essential for the body to function normally. In humans, this is achieved by a complex network of interactions between several organs that is mediated by

hormones and neuropeptide molecules. After a meal, glucose levels rise, and β -cells in the pancreas release the hormone insulin. Insulin binds to its receptor on the surface of liver, muscle and fat cells, stimulating them to take up glucose from the bloodstream, and thereby returning blood glucose levels to normal. Conversely, if blood glucose levels drop, for example between meals, pancreatic α -cells release the hormone glucagon, which stimulates the liver to break down its reserves of a glucose polymer called glycogen. The release of this glucose into the bloodstream once again returns blood glucose levels to normal.

Disturbance in these interplays can cause severe metabolic disorders, such as diabetes, which results in high blood-glucose levels and a range of deleterious effects. In type I diabetes, β -cells are destroyed by immune cells, and insulin is not produced. In type II diabetes, insulin is produced, but its levels are too low



Figure 1 | A laboratory-raised blind Mexican cavefish (*Astyanax mexicanus*) from the Tinaja cave population. Riddle *et al.*¹ report that cavefish from the Tinaja and Pachón caves in Mexico have high blood-glucose levels owing to a mutation in the insulin receptor protein that is identical to one that causes diabetes and a range of health problems in humans^{5,6}. Surprisingly, these cavefish are healthy and have a normal lifespan.

or it is not recognized by the target tissues. This lack of an insulin response is known as insulin resistance.

Although studies of diabetes have focused mainly on mammalian systems, other creatures might offer a fresh perspective on the condition. Cavefish certainly look out of the ordinary, with their depigmented pink skin and eyeless faces (Fig. 1). Hints that these fish have unusual metabolic adaptations for coping with their lightless and seemingly food-poor habitat led Riddle and colleagues to investigate them further. Cavefish have evolved behavioural and physiological traits that differ from those of the river-dwelling members of their species. Compared with river dwellers, cavefish are fatter, owing to a larger lipid reserve²; they are better able to cope with starvation²; have a lower oxygen consumption and metabolic rate³; and the absence of circadian rhythms in their metabolism enables them to make a 27% saving in energy usage⁴.

Riddle and colleagues now reveal another metabolic surprise. Cavefish have higher blood glucose levels than the river dwellers of the species. The authors' investigation revealed that cavefish are insulin resistant and have a mutant version of the insulin receptor, encoded by the *insra* gene, that does not bind and transduce the insulin signal. This mutation, in which the usual amino-acid residue proline is replaced by leucine, is identical to a change that causes diabetes and health problems in humans^{5,6}.

To investigate the effect of this mutation in the fish, the authors used the CRISPR–Cas gene-editing technique to generate zebrafish (*Danio rerio*) carrying mutated *insra*. This experiment would perhaps have been even more informative if performed in

a river-dwelling *A. mexicanus*, but the results are clear: if both the maternally and paternally inherited copy of *insra* have the mutation, the fish are insulin resistant and overweight. The insulin resistance makes sense, given what is known about this mutation in humans, but the weight gain is puzzling and unexpected, because insulin is a growth hormone. In humans and rodents, lack of a functional insulin receptor is associated with growth retardation and low body fat⁷.

Does this mean that insulin signalling has divergent roles in fish and mammals? Zebrafish and *A. mexicanus* cavefish and river dwellers have two versions of the insulin receptor gene, called *insra* and *insrb*. This raises the question of why the receptor encoded by *insrb* doesn't compensate for the non-functional receptor encoded by the mutant version of *insra*. In zebrafish, the two receptors have overlapping roles in growth control and glucose metabolism through the action of the insulin signalling pathway, but they have opposing effects on lipid metabolism, and the receptor encoded by *insra* promotes lipid synthesis^{8,9}. However, in cavefish and in the gene-edited zebrafish, the *insra* gene is mutated, making this weight gain even more puzzling.

Perhaps one of the strangest findings is that cavefish have a respectable lifespan of more than 14 years, which is similar to that of the river-dwelling form. Moreover, they show fewer signs of age-related tissue decline. Elevated blood-glucose levels can damage tissue if the glucose bonds with proteins in a process termed glycation, which is linked to a range of health problems in people who have diabetes. However, cavefish do not have high levels of glycated proteins, so they must have

evolved a compensatory mechanism. One way to unmask this mechanism might be to study cavefish in which gene editing is used to correct the *insra* gene mutation. The authors will undoubtedly be working on this, given the potential biomedical importance of identifying a glycation-inhibiting mechanism.

Further genetic analysis by the authors reveals the complexity of the mechanism that underlies the elevated-glucose phenomenon. The beauty of studying *A. mexicanus* is that the cavefish and river dwellers can interbreed. The authors therefore studied the patterns of inheritance of the elevated-glucose trait, and found that *insra* is not the only gene involved. In the offspring of interbreeding crosses, all of the fish that had high blood glucose had the cavefish *insra* mutation. However, some of the fish with this mutation did not have high blood glucose, indicating that the *insra* mutation is necessary but not sufficient to alter blood glucose regulation. The identification of other genes involved might shed light on the weight-gain phenomenon.

About 30 caves in Mexico are home to cavefish colonies¹⁰. Some of these populations arose from independent colonization events by river-dwelling ancestors. The authors compared three cavefish populations — two from the Tinaja and Pachón caves, which are in the same mountain range, and the third from the Molino cave in another mountain range. The authors found that all of these populations have high blood-glucose levels, but only those from Tinaja and Pachón have the *insra* mutation. Another mechanism must be at work in fish from Molino. The authors' analyses suggest that the insulin signalling pathway is unaffected in these fish, which also have high levels of glycated proteins. Molino cavefish have lived for more than 13 years in our laboratory, suggesting that they also cope well with their unusual metabolism.

The convergent evolution of this extraordinary blood-glucose dysregulation, through different mechanisms in fish from different caves, seems counter-intuitive in terms of species fitness. What compensates for this surprising evolutionary path? The authors suggest that high fat reserves might help the fish to survive during extended periods of food deprivation.

However, perhaps something else is at play. Caves are probably not as food-poor as they seem. In the Tinaja cave (home of the main population investigated by Riddle and colleagues), the cavefish feed on mud that has a high carbon content¹¹. Moreover, a substantial proportion of the fish in the Tinaja and Pachón populations have only one copy of the mutant version of *insra*. If a beneficial mutation has not spread throughout an entire population, one possible interpretation is that the mutation arose relatively recently in the population's history. This would imply that the early cave settlers did not have the benefit of being

fat and starvation-resistant. Exciting times lie ahead for explorations of the metabolic and adaptive changes that occurred as different cavefish populations evolved, and such studies might uncover the underlying evolutionary forces responsible for this striking metabolic adaptation. ■

Sylvie Rétaux is at the Paris-Saclay Institute of Neuroscience, University of Paris-Sud and University of Paris-Saclay, Gif-sur-Yvette 91198, France.
e-mail: retaux@inaf.cnrs-gif.fr

1. Riddle, M. R. *et al. Nature* **555**, 647–651 (2018).
2. Aspiras, A. C., Rohner, N., Martineau, B.,

- Borowsky, R. L. & Tabin, C. J. *Proc. Natl Acad. Sci. USA* **112**, 9668–9673 (2015).
3. Hüppop, K. *Environ. Biol. Fishes* **17**, 229–308 (1986).
4. Moran, D., Softley, R. & Warrant, E. J. *PLoS ONE* **9**, e107877 (2014).
5. Atray, A. *et al. J. Diabetol.* **4** (2), 2 (2013).
6. Carrera, P. *et al. Hum. Mol. Genet.* **2**, 1437–1441 (1993).
7. Savage, D. B. & Semple, R. K. *Curr. Opin. Lipidol.* **21**, 329–336 (2010).
8. Toyoshima, Y. *et al. Endocrinology* **149**, 5996–6005 (2008).
9. Yang, B.-Y. *et al. Am. J. Physiol. Endocrinol. Metab.* <http://dx.doi.org/10.1152/ajpendo.00227.2017> (2017).
10. Mitchell, R. W., Russell, W. H. & Elliott, W. R. *Spec. Publ. Mus. Texas Tech. Univ.* **12** (1977).
11. Simon, V. *et al. EvoDevo* **8**, 23 (2017).

This article was published online on 21 March 2018.

CARDIOLOGY

The tornadoes of sudden cardiac arrest

A clever combination of techniques has enabled, for the first time, simultaneous visualization of the 3D waves of electrical and mechanical activity that are responsible for many cases of sudden cardiac death. [SEE LETTER P.667](#)

JOSÉ JALIFE

Sudden cardiac arrest is a common cause of death in people with coronary artery disease¹, and also kills many young people who have heritable heart diseases. In both cases, sudden death often occurs because of a heart-rhythm defect called ventricular fibrillation^{2,3}, in which the heart rate increases dramatically and cardiac-muscle contractions in the heart's ventricular chambers become uncoordinated, rendering the heart unable to pump blood. Blood pressure decreases, leading to unconsciousness, and death follows unless a defibrillating electrical shock is applied. The mechanism underlying this anomaly has been debated for more than a century. On page 667, Christoph *et al.*⁴ show that the use of panoramic fluorescence imaging with a voltage-sensitive dye, combined with ultrasound imaging, could considerably advance our understanding of ventricular fibrillation. The study has potential implications for a broad range of researchers, from physicists to cardiologists.

Normal heart rhythm is maintained by cyclic changes in the electrical currents that drive and coordinate heart-muscle contraction. But during ventricular fibrillation, electrical impulses stop tracking their normal paths across the heart, and instead adopt a complex, vortex-like pattern similar to eddies in water or a tornado in the atmosphere. Fluorescence imaging of the voltage on the heart's surfaces has revealed⁵ that these electrical

vortices involve highly periodic, spiralling waves of activity organized around a central point called a rotor. The rotor, which is the organizing centre of fibrillation⁶, spins rapidly and can meander across the heart's surface, generating turbulent, wave-like behaviour.

Until now, fluorescence imaging has allowed the visualization of rotors only on the surfaces of the heart's chambers, and so activity inside the ventricles has been inferred through computer simulations⁷. The 3D equivalent of 2D spiral waves (called scroll waves) and the filament-shaped rotors at their centres (vortex filaments) could only be reconstructed using simultaneous 2D video images of spiral waves on the outer and inner surfaces of the heart⁸. This is problematic, because it has not been possible to prove that the electromechanical changes that occur in the heart proper during ventricular fibrillation match predictions made by simulations. For example, cardiac defibrillation by an electrical shock is the only known treatment for ventricular fibrillation, but our understanding of how it works is currently based on numerical predictions.

Christoph *et al.* have overcome this hurdle, thanks to a clever combination of techniques. The authors kept isolated pig and rabbit hearts alive by perfusion with a warm solution through the coronary arteries. They induced ventricular fibrillation in the hearts, and used high-resolution 4D ultrasound imaging to document changing mechanical strain over time.

The group then combined these data with more-conventional 2D fluorescence imaging,



50 Years Ago

Throughout this week an unusual meeting has been taking place near Paris ... experts from all over France — and some from other parts of the world — have been discussing what the world will be like in the year 2020. To keep the discussion within sensible limits, it has been set in a geographical context so that each day has been spent discussing the way in which different technological developments will affect the way land is used. It is clear that this meeting ... is intended to be the first of a series which could, no doubt, last long enough for the results of this meeting to be compared with the reality of 2020.

From *Nature* 30 March 1968

100 Years Ago

Another Indian “miracle” has been explained by scientific investigation. The *Pioneer Mail* of January 11 reports a lecture by Sir J. C. Bose on “The Praying Palm Tree” of Faridpur. While the temple bells call the people to evening prayer, this tree has recently been seen to bow down in prostration, and to erect its head on the following morning. Large numbers of pilgrims have been attracted to the place, and offerings to the tree are said to have been the means of effecting marvellous cures. Sir J. C. Bose first procured photographs which proved the phenomenon to be real. The next step was to devise a special apparatus to record continuously the movement of the tree by day and night. The records showed that it fell with the rise of temperature and rose with the fall. The records obtained in the case of other trees brought out the fact that all the trees are moving, each movement being due to changes in their environment.

From *Nature* 28 March 1918

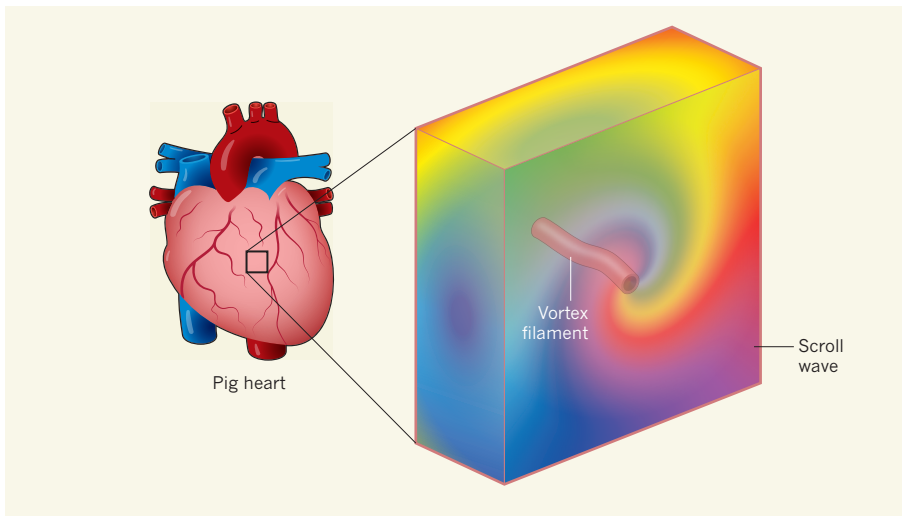


Figure 1 | Electromechanical vortices in a pig heart. In a deadly heart-rhythm defect called ventricular fibrillation, the electrical impulses that normally coordinate heartbeats become erratic. Christoph *et al.*⁴ visualized these erratic electrical impulses, and the corresponding changes in mechanical strain, in 3D in isolated rabbit (not shown) and pig hearts. In line with previous predictions based on 2D data, the authors observed coexisting electrical and mechanical vortex filaments — areas of constant activity around which 3D ‘scroll waves’ of electromechanical activity swirl in rapid, highly periodic cycles (differences in activity indicated in different colours). The scroll wave shown here is a simplified schematic based on Figure 2 of the paper⁴.

in which they took simultaneous measurements of three factors during each heartbeat: changes in electrical potential across cell membranes; changes in intracellular calcium-ion (Ca^{2+}) levels; and muscle contraction. The first two factors provide information about the molecular coupling between electrical and mechanical activity. Because this type of imaging can easily be distorted by spurious mechanical movements, Christoph *et al.* also took the crucial step of using computational motion-tracking techniques to correct any errors made during imaging. Together, these data enabled them to generate 3D maps of how mechanical and electrical waves propagate through the heart during ventricular fibrillation (Fig. 1).

Next, the authors generated computer simulations of electromechanical wave dynamics in virtual hearts that had realistic heart geometry based on computerized-tomography scans and information about muscle-fibre orientation, electrical impulses, electromechanical coupling and muscle contraction. From these simulations, they resolved more details of the dynamics of 3D scroll waves inside the contracting muscle. They found that coexisting electrical and mechanical vortex filaments occurred from the innermost to the outermost linings of the heart, and displayed complex patterns of intrinsically coupled electrical and mechanical waves that swirled at a high frequency through both ventricles.

Christoph and colleagues’ findings are noteworthy because they confirm long-held predictions about the electromechanical behaviour of the heart during ventricular fibrillation. This first experimental observation of electrical and mechanical vortex filaments and

their coupled dynamics also provides proof of a concept that began to emerge more than 60 years ago^{9–14} — that 3D, vortex-like phenomena that occur in certain chemical reactions *in vitro* also cause fibrillation *in vivo*.

When this concept first began to emerge, it opened up a completely new field of cardiac electrophysiology^{12–14}, which is rapidly expanding today. Christoph and colleagues’ study has the potential to contribute to this rapid advance. The tools that the authors used to identify mechanical vortex filaments are readily available in most cardiology units. They could theoretically be used to advance our understanding of cardiac defibrillation and possibly aid the development of pharmacological therapies for ventricular fibrillation. Before long, the authors’ method might be used to monitor fibrillating human hearts from the body’s surface, enabling investigation of the mechanisms underlying ventricular arrhythmias in patients who have structurally normal hearts — such as young people who have heritable heart diseases. However, more animal studies and further technical developments are likely to be needed to translate this strategy to people whose hearts have been scarred by a heart attack¹.

More work will also be required before the imaging technology can be applied to fibrillation in the heart’s atrial chambers. Atrial fibrillation, like ventricular fibrillation, is maintained by high-frequency rotors, and it is the most common cardiac arrhythmia in humans. Unfortunately, currently available ultrasound probes lack sufficient temporal and spatial resolution to study atrial arrhythmias. Christoph and colleagues’ study should

stimulate further advancement of technologies that open the path towards the study and diagnosis of atrial fibrillation in patients.

Another caveat to clinical translation, acknowledged by Christoph and co-workers, is that the generation of mechanical waves requires transient release of Ca^{2+} from intracellular stores into the cell cytoplasm each time the cell receives electrical excitation. The amplitude of the Ca^{2+} transient, and of the mechanical contraction generated by it, depends on the rate of cell excitation. Ca^{2+} transients are readily observed at relatively low frequencies of excitation. But sometimes in ventricular fibrillation, the frequency is too high for Ca^{2+} release. In these cases, Ca^{2+} transients might disappear, or at least diminish substantially¹⁵. Consequently, mechanical waves might also diminish or disappear. This would result in dissociation between the electrical and mechanical waves, which, in turn, would impair the visualization of electromechanical wave dynamics inside the heart wall, because (except in very invasive experimental conditions) electrical waves can be visualized only from the surface.

Despite these limitations, imaging of electromechanical vortex filaments during cardiac fibrillation is an exciting technological accomplishment. Christoph *et al.* have observed, for the first time, complex interactions between electrical and mechanical rotors in 3D, and characterized their dynamics in detail. This important contribution is likely to lead to both scientific and therapeutic advances. ■

José Jalife is in the Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA, and the National Center for Cardiovascular Research Carlos III (CNIC), 28029 Madrid, Spain. He is also at the Center for Biomedical Research in Cardiovascular Diseases Network (CIBERCV), Madrid.
e-mail: jjalife@med.umich.edu

1. Chugh, S. S. *et al.* *Prog. Cardiovasc. Dis.* **51**, 213–228 (2008).
2. Lerma, C. & Glass, L. *J. Physiol. (Lond.)* **594**, 2445–2458 (2016).
3. Mitrani, R. D. & Myerburg, R. J. *Trends Cardiovasc. Med.* **26**, 23–33 (2016).
4. Christoph, J. *et al.* *Nature* **555**, 667–672 (2018).
5. Davidenko, J. M., Pertsov, A. V., Salomonsz, R., Baxter, W. & Jalife, J. *Nature* **355**, 349–351 (1992).
6. Gray, R. A., Pertsov, A. M. & Jalife, J. *Nature* **392**, 75–78 (1998).
7. Baxter, W. T., Mironov, S. F., Zaitsev, A. V., Jalife, J. & Pertsov, A. M. *Biophys. J.* **80**, 516–530 (2001).
8. Yamazaki, M. *et al.* *Cardiovasc. Res.* **94**, 48–57 (2012).
9. Zhabotinsky, A. M. *Biofizika* **9**, 306–311 (1964).
10. Winfree, A. T. *Science* **175**, 634–636 (1972).
11. Winfree, A. T. *Science* **181**, 937–939 (1973).
12. Winfree, A. T. *Sci. Am.* **248**, 144–149 (1983).
13. Winfree, A. T. *Sci. Am.* **248**, 152–157 (1983).
14. Winfree, A. T. *Sci. Am.* **248**, 160–161 (1983).
15. Wang, L. *et al.* *Circ. Res.* **114**, 1410–1421 (2014).

This article was published online on 21 February 2018.

A coherent spin–photon interface in silicon

X. Mi¹, M. Benito², S. Putz¹, D. M. Zajac¹, J. M. Taylor³, Guido Burkard² & J. R. Petta¹

Electron spins in silicon quantum dots are attractive systems for quantum computing owing to their long coherence times and the promise of rapid scaling of the number of dots in a system using semiconductor fabrication techniques. Although nearest-neighbour exchange coupling of two spins has been demonstrated, the interaction of spins via microwave-frequency photons could enable long-distance spin–spin coupling and connections between arbitrary pairs of qubits ('all-to-all' connectivity) in a spin-based quantum processor. Realizing coherent spin–photon coupling is challenging because of the small magnetic-dipole moment of a single spin, which limits magnetic-dipole coupling rates to less than 1 kilohertz. Here we demonstrate strong coupling between a single spin in silicon and a single microwave-frequency photon, with spin–photon coupling rates of more than 10 megahertz. The mechanism that enables the coherent spin–photon interactions is based on spin–charge hybridization in the presence of a magnetic-field gradient. In addition to spin–photon coupling, we demonstrate coherent control and dispersive readout of a single spin. These results open up a direct path to entangling single spins using microwave-frequency photons.

Solid-state electron spins and nuclear spins are quantum mechanical systems that can be almost completely isolated from environmental noise. As a result, they have coherence times as long as hours and so are one of the most promising types of quantum bit (qubit) for constructing a quantum processor^{1–3}. On the other hand, this degree of isolation poses difficulties for the spin–spin interactions that are needed to implement two-qubit gates. So far, most approaches have focused on achieving spin–spin coupling through the exchange interaction or the much weaker dipole–dipole interaction^{4–6}. Among existing classes of spin qubits, electron spins in gate-defined silicon quantum dots have the advantages of scalability due to mature fabrication technologies and low dephasing rates due to isotopic purification⁷. Currently, silicon quantum dots are capable of supporting fault-tolerant control fidelities for single-qubit gates and high-fidelity two-qubit gates based on exchange^{8–12}. Coupling of spins over long distances has been pursued through the physical displacement of electrons^{13–16} and through 'super-exchange' via an intermediate quantum dot¹⁷. The recent demonstration of strong coupling between the charge state of a quantum-dot electron and a single photon has raised the prospect of strong spin–photon coupling, which could enable photon-mediated long-distance spin entanglement^{18–20}. Spin–photon coupling may be achieved by coherently hybridizing spin qubits with photons trapped inside microwave cavities, in a manner similar to cavity quantum electrodynamics with atomic systems and circuit quantum electrodynamics with solid-state qubits^{19–25}. Such an approach, however, is extremely challenging: the small magnetic moment of a single spin leads to magnetic-dipole coupling rates of 10–150 Hz, which are far too slow compared with electron-spin dephasing rates to enable a coherent spin–photon interface^{25–30}.

Here, we resolve this outstanding challenge by using spin–charge hybridization to couple the electric field of a single photon to a single spin in silicon^{25,31–34}. We measure spin–photon coupling rates $g_s/(2\pi)$ of up to 11 MHz, nearly five orders of magnitude higher than typical magnetic-dipole coupling rates. These values of $g_s/(2\pi)$ exceed both the photon decay rate $\kappa/(2\pi)$ and the spin decoherence rate $\gamma_s/(2\pi)$,

firmly anchoring our spin–photon system in the strong-coupling regime^{26,29,30}.

Our coupling scheme consists of two stages of quantum-state hybridization. First, a single electron is trapped within a gate-defined silicon double quantum dot (DQD) that has a large electric-dipole moment. A single photon confined within a microwave cavity hybridizes with the electron charge state through the electric-dipole interaction^{35,36}. Second, a micrometre-scale magnet (micromagnet) placed over the DQD hybridizes electron charge and spin by producing an inhomogeneous magnetic field^{31–34}. The combination of the electric-dipole interaction and spin–charge hybridization gives rise to a large effective spin–photon coupling rate. At the same time, the relatively low level of charge noise in the device ensures that the effective spin decoherence rate γ_s remains below the coherent coupling rate g_s —a criterion that has hampered previous efforts to achieve strong spin–photon coupling³⁷.

As well as demonstrating a coherent spin–photon interface, we also show that our device architecture is capable of single-spin control and readout. Single-spin rotations are electrically driven^{9,38} and the resulting spin state is detected through a dispersive phase shift in the cavity transmission, which reveals Rabi oscillations³⁶.

Spin–photon interface

The device that enables strong spin–photon coupling is shown in Fig. 1a and contains two gate-defined DQDs fabricated using an overlapping aluminium gate stack (Fig. 1b). The gates are electrically biased to create a double-well potential that confines a single electron in the underlying natural-silicon quantum well (Fig. 1c). A plunger gate (P2) on each DQD is connected to the centre pin of a half-wavelength niobium superconducting cavity with a centre frequency of $f_c = 5.846$ GHz and quality factor of $Q_c = 4,700$ ($\kappa/(2\pi) = f_c/Q_c = 1.3$ MHz), which hybridizes the electron charge state with a single cavity photon through the electric-dipole interaction^{18–20,35,36}. Because the spin–photon coupling rate g_s is directly proportional to the charge–photon coupling rate g_c (refs 25, 31–34, 39–41), we have modified the cavity dimensions

¹Department of Physics, Princeton University, Princeton, New Jersey 08544, USA. ²Department of Physics, University of Konstanz, D-78464 Konstanz, Germany. ³Joint Quantum Institute/NIST, College Park, Maryland 20742, USA.

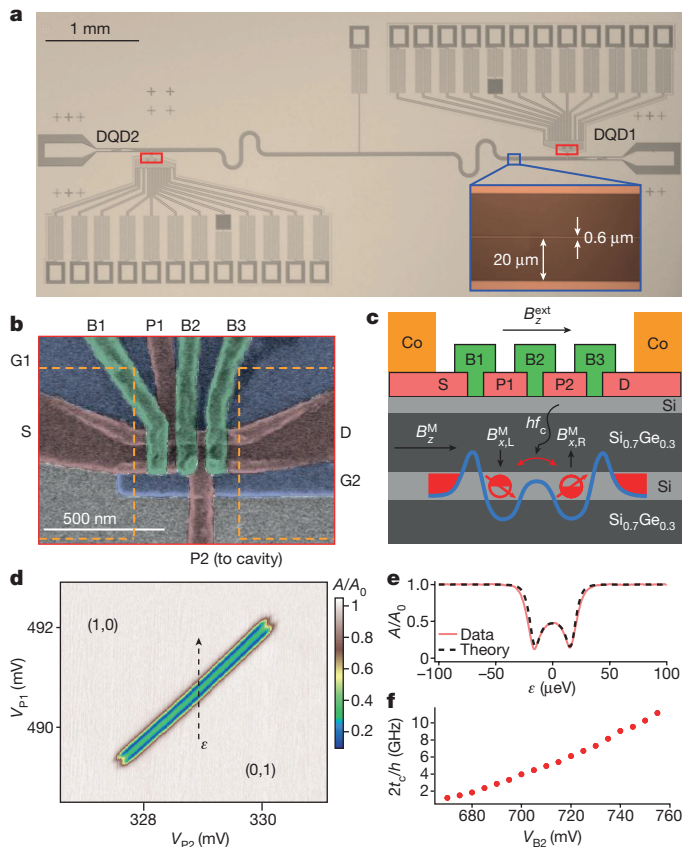


Figure 1 | Spin-photon interface. **a**, Optical image of the superconducting microwave cavity. The inset shows an optical image of the centre pin (0.6 μm) and vacuum gap (20 μm) of the cavity. **b**, False-colour scanning electron micrograph (SEM) of a DQD. Gate electrodes are labelled as G1, G2, S, D, B1, P1, B2, P2 and B3, where G1 and G2 are screening gates, S and D are used for accumulating electrons in the source and drain reservoirs, and B1 and B3 control the tunnel barrier of each dot to its adjacent reservoir. The locations of the cobalt micromagnets are indicated by the orange dashed lines. **c**, Schematic cross-sectional view of the DQD device. The blue line indicates the electrostatic confinement potential which delocalizes a single electron between the two dots (indicated as half-filled circles). The quantization axis of the electron spin (red arrow) changes between the dots. **d**, Cavity transmission amplitude A/A_0 at $f=f_c$, where f_c is the centre frequency of the cavity, near the $(1, 0) \leftrightarrow (0, 1)$ inter-dot transition for DQD1, plotted as a function of the voltages on gates P1 and P2, V_{P1} and V_{P2} , with $B_z^{\text{ext}} = 0$ and $V_{B2} = 710$ mV. The dashed arrow denotes the DQD detuning parameter ε , which is equal to the difference in the chemical potentials of the two dots and points along the vertical direction because in this work V_{P1} is changed to vary ε . V_{B2} denotes the voltage on gate B2, which controls the inter-dot tunnel coupling t_c . **e**, A/A_0 as a function of ε with $V_{B2} = 710$ mV (red line) and a fit to cavity input-output theory (black dashed line), with $g_c/(2\pi) = 40$ MHz. **f**, $2t_c/h$ as a function of V_{B2} for DQD1, obtained by measuring $A(\varepsilon)/A_0$ at different values of V_{B2} .

(Fig. 1a, inset) to achieve a high characteristic impedance Z_r and therefore a high g_c ($g_c \propto \sqrt{Z_r}$; ref. 20). To hybridize the charge state of the trapped electron with its spin state, a cobalt micromagnet is fabricated near the DQD, which generates an inhomogeneous magnetic field. For our device geometry, the magnetic field due to the cobalt micromagnet has a component along the z axis B_z^M that is approximately constant for the DQD and a component along the x axis that takes on an average value of $B_{x,L}^M$ ($B_{x,R}^M$) for the left (right) dot (Fig. 1c, Extended Data Fig. 1). The relatively large field difference $B_{x,R}^M - B_{x,L}^M = 2B_x^M$ leads to spin-charge hybridization, which, when combined with charge-photon coupling, gives rise to spin-photon coupling^{33,34}.

We first characterize the strength of the charge-photon interaction, because this sets the scale of the spin-photon interaction rate.

For simplicity, only one DQD is active at a time for all of the measurements presented here. The cavity is driven by a coherent microwave tone at frequency $f=f_c$ and power $P \approx -133$ dBm (corresponding to approximately 0.6 photons in the cavity, determined on the basis of AC Stark shift measurements of the spin-qubit frequency in the dispersive regime; see Extended Data Fig. 2)⁴². The normalized cavity transmission amplitude A/A_0 is displayed in Fig. 1d as a function of the voltages V_{P1} and V_{P2} on gates P1 and P2 of the first DQD (DQD1), which reveals the location of the $(1, 0) \leftrightarrow (0, 1)$ inter-dot charge transition (see Extended Data Fig. 3 for overall stability diagrams)^{18–20,35,36}. Here (n, m) denotes a charge state, with the number of electrons in the left (P1) and right (P2) dot being n and m , respectively. The charge-photon coupling rate is estimated quantitatively by measuring A/A_0 as a function of the DQD level detuning ε (Fig. 1e). By fitting the data with the cavity input-output theory model using $\kappa/(2\pi) = 1.3$ MHz, we find $g_c/(2\pi) = 40$ MHz and $2t_c/h = 4.9$ GHz, where t_c is the inter-dot tunnel coupling and h is the Planck constant^{19,36,37}. A charge decoherence rate of $\gamma_c/(2\pi) = 35$ MHz is also estimated from the fit and confirmed independently using microwave spectroscopy with $2t_c/h = 5.4$ GHz (refs 19, 20, 42). Fine control of the DQD tunnel coupling, which is critical for achieving spin-charge hybridization³³, is shown in Fig. 1f, in which $2t_c/h$ is plotted as a function of the voltage V_{B2} on the inter-dot barrier gate B2. A similar characterization of the second DQD (DQD2) yields $g_c/(2\pi) = 37$ MHz and $\gamma_c/(2\pi) = 45$ MHz at the $(1, 0) \leftrightarrow (0, 1)$ inter-dot charge transition. Owing to the higher impedance of the resonator, the values of g_c measured here are much larger than in previous silicon DQD devices^{19,43}, which is helpful for achieving strong spin-photon coupling. In general, there are device-to-device variations in γ_c (refs 19, 43). It is unlikely the slightly higher charge decoherence rate is a result of our cavity design, because the Purcell decay rate²⁹ is estimated to be $\Gamma_c/(2\pi) \approx 0.02$ MHz $\ll \gamma_c/(2\pi)$. Excited valley states are not visible in the cavity response of either DQD, suggesting that they have negligible population⁴⁴. We therefore exclude valleys from the analysis below.

Strong single spin-photon coupling

We now demonstrate strong coupling between a single electron spin and a single photon, as evidenced by the observation of vacuum Rabi splitting. Vacuum Rabi splitting occurs when the transition frequency of a two-level atom f_a is brought into resonance with a cavity photon of frequency f_c (refs 21, 23). Light-matter hybridization leads to two vacuum-Rabi-split peaks in the cavity transmission. For our single-spin qubit, the transition frequency between two Zeeman-split spin states is $f_a \approx E_Z/h$, where $E_Z = g\mu_B B_{\text{tot}}$ is the Zeeman energy and the approximate sign is due to spin-charge hybridization, which shifts the qubit frequency slightly. Here g is the g -factor of the electron, μ_B is the Bohr magneton and $B_{\text{tot}} = \sqrt{[(B_{x,L}^M + B_{x,R}^M)/2]^2 + (B_z^M + B_z^{\text{ext}})^2}$ is the total magnetic field. To bring f_a into resonance with f_c , we vary the external magnetic field B_z^{ext} along the z axis while measuring the cavity transmission spectrum A/A_0 as a function of the drive frequency f (Fig. 2a). Vacuum Rabi splittings are clearly observed at $B_z^{\text{ext}} = -91.2$ mT and $B_z^{\text{ext}} = 92.2$ mT, indicating that $E_Z/h = f_c$ at these field values and that the single spin is coherently hybridized with a single cavity photon. These measurements are performed on DQD1, with $2t_c/h = 7.4$ GHz and $\varepsilon = 0$. The dependence of g_s on ε and t_c is investigated below⁴¹. Assuming $g = 2$ for silicon, we estimate that an intrinsic field of about 120 mT is added by the micromagnet, comparable to values found in a previous experiment using a similar cobalt micromagnet design⁹.

To further verify the strong spin-photon coupling, we plot the cavity transmission spectrum at $B_z^{\text{ext}} = 92.2$ mT (Fig. 2b). The two normal-mode peaks are separated by the vacuum Rabi frequency $2g_s/(2\pi) = 11.0$ MHz, giving an effective spin-photon coupling rate of $g_s/(2\pi) = 5.5$ MHz. The photon decay rate at finite magnetic field is extracted from the line width of A/A_0 at $B_z^{\text{ext}} = 90.3$ mT, at which E_Z/h is largely detuned from f_c , yielding $\kappa/(2\pi) = 1.8$ MHz. A spin decoherence rate of $\gamma_s/(2\pi) = 2.4$ MHz, with contributions from both charge

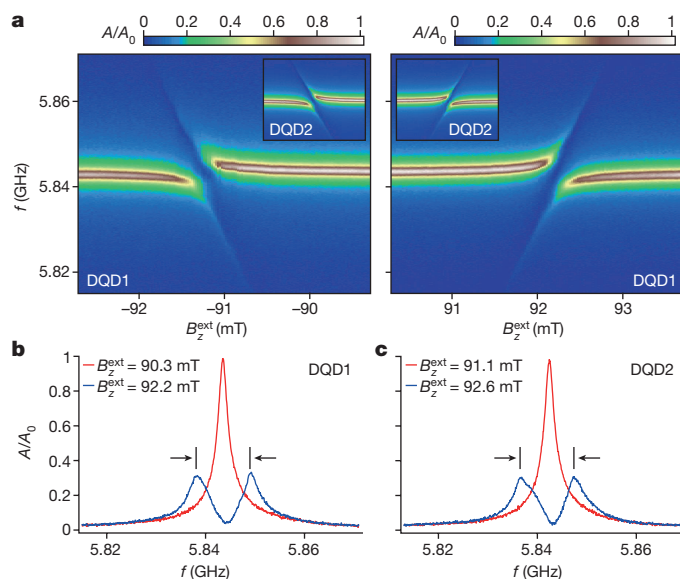


Figure 2 | Strong single spin-photon coupling. **a**, A/A_0 as a function of the cavity drive frequency f and an externally applied magnetic field B_z^{ext} for DQD1. Insets show data from DQD2 at the same values of t_c and ε and plotted over the same range of f . B_z^{ext} ranges from -94 mT to -91.1 mT (91.1 mT to 94 mT) for the left (right) inset. **b**, A/A_0 as a function of f for DQD1 at $B_z^{\text{ext}} = 90.3$ mT (red) and $B_z^{\text{ext}} = 92.2$ mT (blue). **c**, A/A_0 as a function of f for DQD2 at $B_z^{\text{ext}} = 91.1$ mT (red) and $B_z^{\text{ext}} = 92.6$ mT (blue). In **b** and **c**, the frequency difference between the two transmission peaks, indicated by the black arrows, is 11.0 MHz (**b**) and 10.6 MHz (**c**). The spin-photon coupling rate $g_s/(2\pi)$ corresponds to half the frequency separation and so is 5.5 MHz for DQD1 and 5.3 MHz for DQD2.

decoherence and magnetic noise from the ^{29}Si nuclei, is extracted from microwave spectroscopy in the dispersive regime with $2t_c/h = 7.4$ GHz and $\varepsilon = 0$ (Extended Data Fig. 4), confirming that the strong-coupling regime $g_s > \gamma_s$, κ has been reached. The spin-photon coupling rate obtained here is more than four orders of magnitude larger than rates currently achievable using direct magnetic-dipole coupling to lumped-element superconducting resonators^{30,45}.

The local magnetic field that is generated using cobalt micromagnets is very reproducible, as evidenced by examining the other DQD in the cavity. Measurements on DQD2 show vacuum Rabi splittings at $B_z^{\text{ext}} = \pm 92.6$ mT (Fig. 2a, insets). The spin-photon coupling rate and spin decoherence rate are determined to be $g_s/(2\pi) = 5.3$ MHz and $\gamma_s/(2\pi) = 2.4$ MHz, respectively (Fig. 2c). These results are highly consistent with DQD1, and so we henceforth focus on DQD1.

Electrical control of spin-photon coupling

For quantum information applications it is desirable to turn qubit-cavity coupling rapidly on for quantum-state transfer and rapidly off for qubit-state preparation. Rapid control of the coupling rate is often accomplished by quickly modifying the qubit-cavity detuning $f_a - f_c$. Practically, such tuning can be achieved by varying the qubit transition frequency f_a with voltage or flux pulses^{46,47} or by using a tunable cavity²⁰. These approaches are not directly applicable for control of the spin-photon coupling rate because f_a depends primarily on magnetic fields that are difficult to vary on nanosecond timescales. In this section, we show that control of the spin-photon coupling rate can be achieved electrically by tuning ε and t_c (refs 32, 40).

We first investigate the ε dependence of g_s . In Fig. 3a we show measurements of A/A_0 as a function of B_z^{ext} and f for $\varepsilon = 0$, $\varepsilon = 20$ μeV and $\varepsilon = 40$ μeV . At $\varepsilon = 20$ μeV (about 4.8 GHz), vacuum Rabi splitting is observed at $B_z^{\text{ext}} = 92.1$ mT with a spin-photon coupling rate of $g_s/(2\pi) = 1.0$ MHz that is substantially lower than the value of $g_s/(2\pi) = 5.5$ MHz obtained at $\varepsilon = 0$. At $\varepsilon = 40$ μeV (about 9.7 GHz), only a small dispersive shift is observed in the cavity transmission

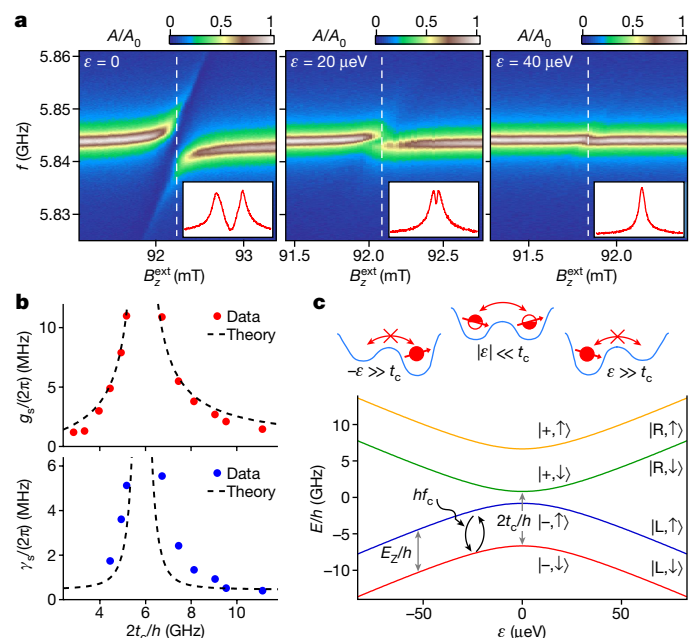


Figure 3 | Electrical control of spin-photon coupling. **a**, A/A_0 as a function of f and B_z^{ext} at $\varepsilon = 0$ (left), $\varepsilon = 20$ μeV (about 4.8 GHz; middle) and $\varepsilon = 40$ μeV (about 9.7 GHz; right), with $2t_c/h = 7.4$ GHz. Insets show A/A_0 as a function of f at the values of B_z^{ext} indicated by the white dashed lines in the main panels. **b**, Spin-photon coupling rate $g_s/(2\pi)$ (top) and spin decoherence rate $\gamma_s/(2\pi)$ (bottom) as functions of $2t_c/h$, with $\varepsilon = 0$ (data). The dashed lines show theoretical predictions. A potential uncertainty of 0.01 – 0.1 MHz exists for each value of $g_s/(2\pi)$ and $\gamma_s/(2\pi)$ owing to uncertainties in the locations of the transmission peaks used to determine $g_s/(2\pi)$ (Extended Data Fig. 5) and the widths of the Lorentzian fits used to determine $\gamma_s/(2\pi)$ (Extended Data Fig. 4). **c**, DQD energy levels as a function of ε , calculated with $B_z^{\text{ext}} + B_z^{\text{M}} = 209$ mT, $B_z^{\text{M}} = 15$ mT and $2t_c/h = 7.4$ GHz. Here B_z^{M} denotes the magnetic field produced by the cobalt micromagnet parallel to B_z^{ext} , and B_x^{M} is related to the strength of the inhomogeneous magnetic field perpendicular to B_z^{ext} . The symbols \uparrow (\downarrow), L (R) and $-$ ($+$) denote the quantum states of the electron that correspond to up (down) spin states, left-dot (right-dot) orbital states and molecular bonding (anti-bonding) states, respectively. The schematics at the top illustrate the distribution of the wavefunction of the electron in different regimes of ε . For $\varepsilon \gg t_c$ and $-\varepsilon \gg t_c$, the electron is localized within one dot and tunnelling between the dots is largely forbidden, resulting in a small g_s due to a small effective oscillating magnetic field. For $|\varepsilon| \ll t_c$, the electron may tunnel between the two dots and experience a large oscillating magnetic field due to the spatial field gradient, resulting in a large g_s .

spectrum at $B_z^{\text{ext}} = 91.8$ mT, suggesting a further decrease in g_s . These observations are qualitatively understood by considering that at $\varepsilon = 0$ the electron is delocalized across the DQD and forms molecular bonding ($|-\rangle$) or anti-bonding ($|+\rangle$) charge states (Fig. 3c). In this regime, the cavity electric field leads to a displacement of the electron wavefunction of the order of 1 nm (Methods)³³. Consequently, the electron spin experiences a large oscillating magnetic field, resulting in a substantial spin-photon coupling rate. By contrast, with $|\varepsilon| \gg t_c$ the electron is localized within one dot and it is natural to work with a basis of localized electronic wavefunctions $|\text{L}\rangle$ and $|\text{R}\rangle$, where L and R correspond to the electron being in the left and right dot, respectively (Fig. 3c). In this effectively single-dot regime, the displacement of the electron wavefunction by the cavity electric field is estimated to be about 3 pm for a single-dot orbital energy of $E_{\text{orb}} = 2.5$ meV (ref. 48), greatly suppressing the spin-photon coupling mechanism³³. The large difference in the effective displacement lengths between the single-dot and double-dot regimes also implies an improvement in the spin-photon coupling rate at $\varepsilon = 0$ of approximately two orders of magnitude compared to $|\varepsilon| \gg t_c$. Alternatively, the reduction of g_s may be

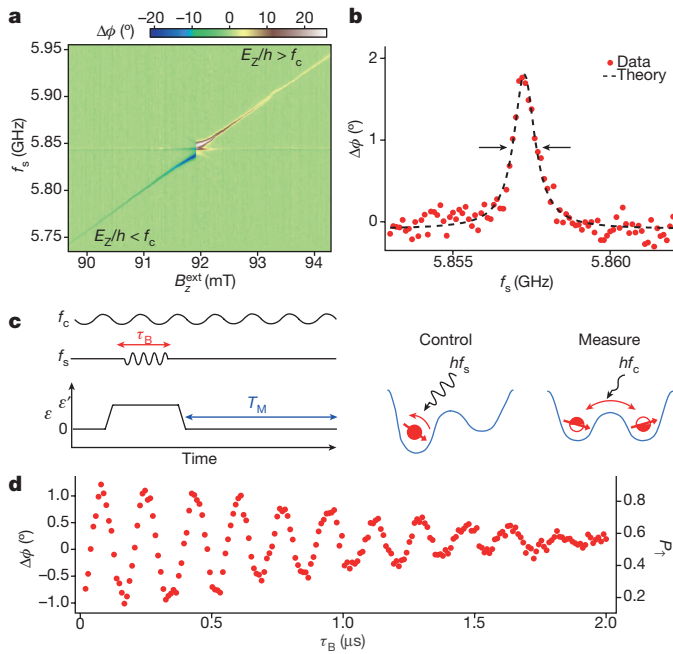


Figure 4 | Quantum control and dispersive readout of a single spin.

a, Cavity phase response $\Delta\phi$ at $f = f_c$ when gate P1 is driven continuously at a variable frequency f_s and power $P_s = -106$ dBm, with $2t_c/h = 9.5$ GHz and $\varepsilon = 0$. A background phase response, obtained by measuring $\Delta\phi(B_z^{\text{ext}})$ in the absence of a microwave drive on P1, is subtracted from each column of the data to correct for slow drifts in the microwave phase. **b**, Electron spin resonance (ESR) line as measured in $\Delta\phi(f_s)$ at $2t_c/h = 11.1$ GHz, $\varepsilon = 0$, $B_z^{\text{ext}} = 92.18$ mT and $P_s = -123$ dBm (data). The dashed line shows a fit to a Lorentzian with a full-width at half-maximum of $\gamma_s/\pi = 0.81 \pm 0.04$ MHz (indicated by the arrows). **c**, Schematic showing the experimental sequence for coherent spin control and measurement. Spin control is performed using a high-power microwave burst when the electron is largely localized within one dot ($|\varepsilon| \gg t_c$) and spin-photon coupling is turned off. Spin-state readout is achieved using the dispersive response of a cavity photon at $\varepsilon = 0$ and when spin-photon coupling is turned on. **d**, $\Delta\phi$ as a function of τ_B , with $2t_c/h = 11.1$ GHz and $\varepsilon' = 70$ μeV, showing single-spin Rabi oscillations. The excited-state population of the spin qubit P_1 is indicated on the right y axis (see Methods).

interpreted as a result of suppressed hybridization between the $|- , \uparrow\rangle$ and $|+ , \downarrow\rangle$ states due to their growing energy difference at larger $|\varepsilon|$, as evident from Fig. 3c (see discussion below). Here \uparrow (\downarrow) denotes an electron spin that is aligned (anti-aligned) with B_z^{ext} . These measurements highlight the important role of charge hybridization in the DQD.

Additional electric control of g_s is enabled by voltage tuning t_c (Fig. 1f). In Fig. 3b we show $g_s/(2\pi)$ and $\gamma_s/(2\pi)$ as functions of $2t_c/h$ at $\varepsilon = 0$, as extracted from vacuum Rabi splitting measurements and microwave spectroscopy of the electron spin resonance (ESR) transition line width (Figs 2b, 4b, Extended Data Figs 4, 5). Both rates increase rapidly as $2t_c/h$ approaches the Larmor precession frequency $E_z/h \approx 5.8$ GHz, and a spin-photon coupling rate as high as $g_s/(2\pi) = 11.0$ MHz is found at $2t_c/h = 5.2$ GHz. These trends are consistent with the DQD energy-level spectrum shown in Fig. 3c^{33,34,41}. With $2t_c/h \gg E_z/h$ and $\varepsilon = 0$, the two lowest energy levels are $|- , \downarrow\rangle$ and $|+ , \uparrow\rangle$ and the electric-dipole coupling to the cavity field is small. As $2t_c$ is reduced and made comparable to E_z , the ground state remains $|- , \downarrow\rangle$ but the excited state becomes an admixture of $|- , \uparrow\rangle$ and $|+ , \downarrow\rangle$ owing to the magnetic-field gradient $B_{x,R}^M - B_{x,L}^M = 2B_x^M$ and the small energy difference between the states. The quantum transition that is close to resonance with E_z is now partially composed of a change in charge state from $-$ to $+$, which responds strongly to the cavity electric field and gives rise to larger values of g_s . For $2t_c/h < E_z/h$, a decrease in t_c increases the energy difference between $|- , \uparrow\rangle$ and $|+ , \downarrow\rangle$, which reduces their hybridization and results in a smaller g_s . We note that hybridization with charge states increases the

susceptibility of the spin to charge noise and relaxation, resulting in an effective spin decoherence rate γ_s that is also strongly dependent on t_c (Fig. 3b)^{33,34,41}. Theoretical predictions of g_s and γ_s as functions of $2t_c/h$, based on measured values of g_c and γ_c (Fig. 1e) are in good agreement with the data (Fig. 3b)⁴¹. The discrepancy in the fit of γ_s is discussed in Methods. The electric control of spin-photon coupling demonstrated here allows the spin qubit to switch quickly between regimes with strong coupling to the cavity and idle regimes in which the spin-photon coupling rate and susceptibility to charge decoherence are small.

Dispersive readout of a single spin

The preceding measurements demonstrate the ability to couple a single electron spin to a single photon coherently, potentially enabling long-range spin-spin couplings^{46,47}. For the device to serve as a building block for a quantum processor, it is also necessary to prepare, control and read out the spin state of the trapped electron deterministically. We first induce spin transitions by driving gate P1 with a continuous microwave tone of frequency f_s and power $P_s = -106$ dBm. When $f_s \approx E_z/h$, the excited-state population of the spin qubit P_1 increases and the ground state-population P_0 decreases. In the dispersive regime, in which the qubit-cavity detuning $\Delta/(2\pi) \approx E_z/h - f_c$ satisfies $|\Delta/(2\pi)| \gg g_s/(2\pi)$, the cavity transmission experiences a phase response $\Delta\phi \approx \tan^{-1}[2g_s^2/(\kappa\Delta)]$ for a fully saturated ($P_1 = 0.5$) qubit^{19,42}. It is therefore possible to measure the spin state of a single electron by probing the cavity transmission. As a demonstration, we spectroscopically probe the ESR transition by measuring $\Delta\phi$ as a function of f_s and B_z^{ext} (Fig. 4a). These data are acquired with $2t_c/h = 9.5$ GHz and $\varepsilon = 0$. The ESR transition is clearly visible as a narrow feature with $\Delta\phi \neq 0$ that shifts to higher f_s with increasing B_z^{ext} . $\Delta\phi$ also changes sign as B_z^{ext} increases, consistent with the sign change of the qubit-cavity detuning Δ when the Larmor precession frequency E_z/h exceeds f_c . The nonlinear response in the small region around $B_z^{\text{ext}} = 92$ mT is due to the breakdown of the dispersive condition $|\Delta/(2\pi)| \gg g_s/(2\pi)$.

Finally, we demonstrate coherent single-spin control and dispersive spin-state readout. For these measurements, we choose $\varepsilon = 0$ and $2t_c/h = 11.1$ GHz to minimize the spin decoherence rate γ_s (Fig. 3b). Here the spin-photon coupling rate $g_s/(2\pi) = 1.4$ MHz (Fig. 3b). The external field is fixed at $B_z^{\text{ext}} = 92.18$ mT, which ensures that the system is in the dispersive regime with $\Delta/(2\pi) = 14$ MHz $\gg g_s/(2\pi)$. A measurement of $\Delta\phi(f_s)$ in the low-power limit (Fig. 4b) yields a Lorentzian line shape with a full-width at half-maximum of 0.81 MHz, which corresponds to a low spin decoherence rate of $\gamma_s/(2\pi) = 0.41$ MHz (refs 19, 42). Qubit control and measurement are achieved using the pulse sequence illustrated in Fig. 4c. Starting with a spin-down state $|\downarrow\rangle$ at $\varepsilon = 0$, the DQD is pulsed to a large detuning $\varepsilon' = 70$ μeV (about 17 GHz), which decouples the spin from the cavity. A microwave burst with frequency $f_s = 5.874$ GHz, power $P_s = -76$ dBm and duration T_B is subsequently applied to P1 to drive a spin rotation^{9,36,38}. The DQD is then pulsed adiabatically back to $\varepsilon = 0$ for a fixed measurement time T_M for dispersive readout. To reinitialize the qubit, we choose $T_M = 20$ μs $\gg T_1(\varepsilon = 0)$, where $T_1(\varepsilon = 0) = 3.2$ μs is the spin relaxation time measured at $\varepsilon = 0$ (Extended Data Fig. 6). Figure 4d displays the time-averaged $\Delta\phi$ as a function of τ_B , obtained with an integration time of 100 ms for each data point. We observe coherent single-spin Rabi oscillations with a Rabi frequency of $f_R = 6$ MHz. In contrast to readout approaches that rely on spin-dependent tunneling^{9,38,49}, our dispersive cavity-based readout corresponds in principle to quantum non-demolition readout²⁴. The readout scheme is also distinct from previous work that used a cavity-coupled InAs DQD, which detects the spin state through Pauli blockade rather than spin-photon coupling³⁶. In addition to enabling single spin-photon coupling, our device is capable of preparing, controlling and dispersively reading out single spins.

Conclusion

We have realized a coherent spin-photon interface at which a single spin in a silicon DQD is strongly coupled to a microwave-frequency

photon through the combined effects of the electric-dipole interaction and spin–charge hybridization (see Methods for a discussion of the prospects of applying the spin–photon interface to realize cavity-mediated spin–spin coupling). Spin–photon coupling rates of up to 11 MHz are measured in the device, exceeding magnetic-dipole coupling rates by nearly five orders of magnitude. The spin decoherence rate is strongly dependent on the inter-dot tunnel coupling t_c and ranges from 0.4 MHz to 6 MHz, possibly limited by a combination of charge noise, charge relaxation and remnant nuclear field fluctuations. All-electric control of spin–photon coupling and coherent manipulation of the spin state are demonstrated, along with dispersive readout of the single spin, which lays the foundation for quantum non-demolition readout of spin qubits. These results could enable the construction of an ultra-coherent spin quantum computer with photonic interconnects and readout channels, with the capacity for surface codes, ‘all-to-all’ connectivity and easy integration with other solid-state quantum systems such as superconducting qubits^{24,46,47,50–52}.

We note that two related preprints appeared after the submission of this Article: ref. 53 presents some of the results discussed here, and ref. 54 explores a different approach to spin–photon coupling and demonstrates coherent coupling of a triple quantum dot to microwave-frequency photons.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 September 2017; accepted 23 January 2018.

Published online 14 February 2018.

1. Tyryshkin, A. M. *et al.* Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2012).
2. Saeedi, K. *et al.* Room-temperature quantum bit storage exceeding 39 minutes using ionized donors in silicon-28. *Science* **342**, 830–833 (2013).
3. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
4. Petta, J. R. *et al.* Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
5. Neumann, P. *et al.* Multiparticle entanglement among single spins in diamond. *Science* **320**, 1326–1329 (2008).
6. Dehollain, J. P. *et al.* Bell’s inequality violation with spins in silicon. *Nat. Nanotechnol.* **11**, 242–246 (2016).
7. Zwanenburg, F. A. *et al.* Silicon quantum electronics. *Rev. Mod. Phys.* **85**, 961–1019 (2013).
8. Veldhorst, M. *et al.* An addressable quantum dot qubit with fault-tolerant control-fidelity. *Nat. Nanotechnol.* **9**, 981–985 (2014).
9. Takeda, K. *et al.* A fault-tolerant addressable spin qubit in a natural silicon quantum dot. *Sci. Adv.* **2**, e1600694 (2016).
10. Veldhorst, M. *et al.* A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015).
11. Zajac, D. M. *et al.* Resonantly driven CNOT gate for electron spins. *Science* **359**, 439–442 (2018).
12. Watson, T. F. *et al.* A programmable two-qubit quantum processor in silicon. *Nature* <https://doi.org/10.1038/nature25766> (2018).
13. McNeil, R. P. G. *et al.* On-demand single-electron transfer between distant quantum dots. *Nature* **477**, 439–442 (2011).
14. Baart, T. A. *et al.* Single-spin CCD. *Nat. Nanotechnol.* **11**, 330–334 (2016).
15. Bertrand, B. *et al.* Fast spin information transfer between distant quantum dots using individual electrons. *Nat. Nanotechnol.* **11**, 672–676 (2016).
16. Flentje, H. *et al.* Coherent long-distance displacement of individual electron spins. *Nat. Commun.* **8**, 501 (2017).
17. Baart, T. A., Fujita, T., Reichl, C., Wegscheider, W. & Vandersypen, L. M. K. Coherent spin-exchange via a quantum mediator. *Nat. Nanotechnol.* **12**, 26–30 (2016).
18. Bruhat, L. E. *et al.* Strong coupling between an electron in a quantum dot circuit and a photon in a cavity. Preprint at <https://arxiv.org/abs/1612.05214> (2016).
19. Mi, X., Cady, J. V., Zajac, D. M., Deelman, P. W. & Petta, J. R. Strong coupling of a single electron in silicon to a microwave photon. *Science* **355**, 156–158 (2017).
20. Stockklauser, A. *et al.* Strong coupling cavity QED with gate-defined double quantum dots enabled by a high impedance resonator. *Phys. Rev. X* **7**, 011030 (2017).
21. Thompson, R. J., Rempe, G. & Kimble, H. J. Observation of normal-mode splitting for an atom in an optical cavity. *Phys. Rev. Lett.* **68**, 1132–1135 (1992).
22. Brune, M. *et al.* Quantum Rabi oscillation: a direct test of field quantization in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).
23. Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
24. Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
25. Childress, L., Sørensen, A. S. & Lukin, M. D. Mesoscopic cavity quantum electrodynamics with quantum dots. *Phys. Rev. A* **69**, 042302 (2004).
26. Imamoglu, A. Cavity QED based on collective magnetic dipole coupling: spin ensembles as hybrid two-level systems. *Phys. Rev. Lett.* **102**, 083602 (2009).
27. Schuster, D. I. *et al.* High-cooperativity coupling of electron-spin ensembles to superconducting cavities. *Phys. Rev. Lett.* **105**, 140501 (2010).
28. Amsüss, R. *et al.* Cavity QED with magnetically coupled collective spin states. *Phys. Rev. Lett.* **107**, 060502 (2011).
29. Bienfait, A. *et al.* Controlling spin relaxation with a cavity. *Nature* **531**, 74–77 (2016).
30. Eichler, C. *et al.* Electron spin resonance at the level of 10^4 spins using low impedance superconducting resonators. *Phys. Rev. Lett.* **118**, 037701 (2017).
31. Trif, M., Golovach, V. N. & Loss, D. Spin dynamics in InAs nanowire quantum dots coupled to a transmission line. *Phys. Rev. B* **77**, 045434 (2008).
32. Cottet, A. & Kontos, T. Spin quantum bit with ferromagnetic contacts for circuit QED. *Phys. Rev. Lett.* **105**, 160502 (2010).
33. Hu, X., Liu, Y.-x. & Nori, F. Strong coupling of a spin qubit to a superconducting stripline cavity. *Phys. Rev. B* **86**, 035314 (2012).
34. Beaudoin, F., Lachance-Quirion, D., Coish, W. A. & Pioro-Ladriere, M. Coupling a single electron spin to a microwave resonator: controlling transverse and longitudinal couplings. *Nanotechnology* **27**, 464003 (2016).
35. Frey, T. *et al.* Dipole coupling of a double quantum dot to a microwave resonator. *Phys. Rev. Lett.* **108**, 046807 (2012).
36. Petersson, K. D. *et al.* Circuit quantum electrodynamics with a spin qubit. *Nature* **490**, 380–383 (2012).
37. Viennot, J. J., Dartiaill, M. C., Cottet, A. & Kontos, T. Coherent coupling of a single spin to microwave cavity photons. *Science* **349**, 408–411 (2015).
38. Kawakami, E. *et al.* Electrical control of a long-lived spin qubit in a Si/SiGe quantum dot. *Nat. Nanotechnol.* **9**, 666–670 (2014).
39. Burkard, G. & Imamoglu, A. Ultra-long-distance interaction between spin qubits. *Phys. Rev. B* **74**, 041307 (2006).
40. Jin, P.-Q., Marthaler, M., Shnirman, A. & Schon, G. Strong coupling of spin qubits to a transmission line resonator. *Phys. Rev. Lett.* **108**, 190506 (2012).
41. Benito, M., Mi, X., Taylor, J. M., Petta, J. R. & Burkard, G. Input-output theory for spin-photon coupling in Si double quantum dots. *Phys. Rev. B* **96**, 235434 (2017).
42. Schuster, D. I. *et al.* ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
43. Mi, X. *et al.* Circuit quantum electrodynamics architecture for gate-defined quantum dots in silicon. *Appl. Phys. Lett.* **110**, 043502 (2017).
44. Mi, X., Peterfalvi, C. G., Burkard, G. & Petta, J. R. High-resolution valley spectroscopy of Si quantum dots. *Phys. Rev. Lett.* **119**, 176803 (2017).
45. Probst, S. *et al.* Inductive-detection electron-spin resonance spectroscopy with 65 spins/Hz sensitivity. *Appl. Phys. Lett.* **111**, 202604 (2017).
46. Majer, J. *et al.* Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
47. Sillanpää, M. A., Park, J. I. & Simmonds, R. W. Coherent quantum state storage and transfer between two phase qubits via a resonant cavity. *Nature* **449**, 438–442 (2007).
48. Zajac, D. M., Hazard, T. M., Mi, X., Wang, K. & Petta, J. R. A reconfigurable gate architecture for Si/SiGe quantum dots. *Appl. Phys. Lett.* **106**, 223507 (2015).
49. Elzerman, J. M. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).
50. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
51. Debnath, S. *et al.* Demonstration of a small programmable quantum computer with atomic qubits. *Nature* **536**, 63–66 (2016).
52. Nigg, S. E., Fuhrer, A. & Loss, D. Superconducting grid-bus surface code architecture for hole-spin qubits. *Phys. Rev. Lett.* **118**, 147701 (2017).
53. Samkharadze, N. *et al.* Strong spin-photon coupling in silicon. Preprint at <https://arxiv.org/abs/1711.02040> (2017).
54. Landig, A. J. *et al.* Coherent spin-qubit photon coupling. Preprint at <https://arxiv.org/abs/1711.01932> (2017).

Acknowledgements We thank A. J. Sigillito for technical assistance and M. J. Gullans for discussions. This work was supported by the US Department of Defense under contract H98230-15-C0453, Army Research Office grant W911NF-15-1-0149, and the Gordon and Betty Moore Foundations EPIQS Initiative through grant GBMF4535. Devices were fabricated in the Princeton University Quantum Device Nanofabrication Laboratory.

Author Contributions X.M. fabricated the sample and performed the measurements. X.M., D.M.Z. and J.R.P. developed the design and fabrication process for the DQD. X.M. and S.P. developed the niobium cavity fabrication process. M.B., G.B., J.M.T. and J.R.P. developed the theory for the experiment. X.M., M.B. and J.M.T. analysed the data. X.M., J.R.P., G.B. and J.M.T. wrote the manuscript with input from the other authors. J.R.P. planned and supervised the experiment.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.R.P. (petta@princeton.edu).

Reviewer Information Nature thanks T. Meunier and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Device fabrication and measurement. The Si/SiGe heterostructure consists of a 4-nm-thick Si cap, a 50-nm-thick Si_{0.7}Ge_{0.3} spacer layer, a 8-nm-thick natural-Si quantum well and a 225-nm-thick Si_{0.7}Ge_{0.3} layer on top of a linearly graded Si_{1-x}Ge_x relaxed buffer substrate. Design and fabrication details for the superconducting cavity and DQDs are described elsewhere⁴³. The approximately 200-nm-thick Co micromagnet is defined using electron beam lithography and lift off. In contrast to earlier devices, the gate filter for P1 was changed to an L_1 - C - L_2 filter, with $L_1 = 4$ nH, $C = 1$ pF and $L_2 = 12$ nH (ref. 43). This three-segment filter allows microwave signals below 2.5 GHz to pass with less than 3 dB of attenuation.

All data are acquired in a dilution refrigerator with a base temperature of 10 mK and electron temperature of $T_e = 60$ mK. The measurements of the transmission amplitude and phase response of the cavity (Figs 1, 4) are performed using a homodyne detection scheme²³. The measurements of the transmission spectra of the cavity (Figs 2, 3) are performed using a network analyser. The microwave drive applied to P1 (Fig. 4) is provided by a vector microwave source and the detuning pulses are generated by an arbitrary waveform generator, which also controls the timing of the microwave burst (Fig. 4d).

To maximize the magnetization of the Co micromagnet and minimize hysteresis, data at positive (negative) external applied magnetic fields (Fig. 2a) are collected after B_z^{ext} is first ramped to a large value of +300 mT (−300 mT). A small degree of hysteresis still remains for the micromagnet of DQD1, as can be seen by the different magnitudes of B_z^{ext} at which positive- and negative-field vacuum Rabi splittings are observed (Fig. 2a). In Fig. 4a, the slope of the ESR transition is $d(E_z/h)/dB_z^{\text{ext}} = 44$ MHz mT^{−1}, which is higher than the value (28 MHz mT^{−1}) expected for a fully saturated micromagnet. The slope of the transition suggests that the micromagnet is not fully magnetized and has a magnetic susceptibility of $dB_z^{\text{ext}}/dB_z^{\text{ext}} \approx 0.6$ around $B_z^{\text{ext}} = 92$ mT.

Estimate of displacement length. The displacement length of the electron wavefunction by the cavity electric field may be estimated by considering the spin-photon coupling strength. For $g_s/(2\pi) = 10$ MHz, the effective AC magnetic field $B_{\text{ac}}^{\text{ESR}}$ that drives ESR is $B_{\text{ac}}^{\text{ESR}} = [g_s/(2\pi)][h/(g\mu_B)] \approx 0.4$ mT. The field gradient for our DQD is $2B_x^M/l \approx 0.3$ mT nm^{−1}, where $l = 100$ nm is the inter-dot distance. Therefore, the effective displacement of the electron wavefunction is estimated to be about 1 nm in the DQD regime. For a single dot, the spin-photon coupling strength is expected to be $g_s/(2\pi) \approx (g\mu_B B_x^M/E_{\text{orb}})(g_c/2\pi) \approx 30$ kHz (refs 31, 33) for an orbital energy of $E_{\text{orb}} = 2.5$ meV (ref. 48). The equivalent AC magnetic field that is induced by the cavity is therefore $B_{\text{ac}}^{\text{ESR}} \approx 1$ μT, corresponding to a displacement length of only about 3 pm.

Conversion of cavity phase response to spin population. For the dispersive readout of the Rabi oscillations (Fig. 4d), the theoretically expected cavity phase response is $\phi_1 = \tan^{-1}[2g_s^2/(\kappa\Delta)] = 9.6^\circ$ when the spin qubit is in the excited state, and $\phi_1 = -\tan^{-1}[2g_s^2/(\kappa\Delta)] = -9.6^\circ$ when the spin qubit is in the ground state^{42,55}. Because our measurement is averaged over $T_M \gg T_1$, spin relaxation during readout will reduce the phase contrast observed in the experiment. To enable a conversion between the phase response of the cavity $\Delta\phi$ and the excited-state population of the spin qubit P_1 , we measure the spin relaxation time T_1 by fixing the microwave burst time at $\tau_B = 80$ ns, which corresponds to a π pulse on the spin qubit. The phase response of the cavity $\Delta\phi$ is then measured as a function of T_M for $T_M > 5 \mu\text{s} > T_1$ (Extended Data Fig. 6). The result is fitted to a function of the form $\Delta\phi = \phi_0 + \phi_1(T_1/T_M)[1 - \exp(-T_M/T_1)]$ to extract $T_1 = 3.2 \mu\text{s}$, where ϕ_0 and ϕ_1 are additional fitting parameters³⁶. We have ignored the effects of the cavity ringdown time $1/\kappa \approx 90$ ns and the π -pulse time of 80 ns in the fit, because both of these times are much shorter than the measurement time T_M . The phase contrast that results from the fit, $\phi_1 \approx 17.7^\circ$, is close to the maximum contrast expected at this spin-photon detuning, $\phi_1 - \phi_1 = 19.2^\circ$. On the basis of this value of T_1 , we convert the measured phase response into the excited-state population via $P_1 = (1 + \Delta\phi/\phi_1)/2$, where $\phi_1 = \phi_1(T_1/T_M)[1 - \exp(-T_M/T_1)] = 1.5^\circ$ is the reduced phase response due to spin relaxation during the readout time $T_M = 20 \mu\text{s}$. The converted spin population P_1 shown in Fig. 4d has a visibility of about 70%, which could be improved by performing single-shot measurements⁵⁵.

Input-output theory for cavity transmission. Here we briefly summarize the theoretical methods used to calculate the cavity transmission A/A_0 shown in Fig. 1e and Extended Data Fig. 7; see ref. 41 for a detailed description of the theory. We start from the Hamiltonian that describes the DQD

$$H_0 = \frac{1}{2}(\varepsilon\tau_z + 2t_c\tau_x + B_z\sigma_z + B_x^M\sigma_x\tau_z) \quad (1)$$

where τ_x and τ_z are Pauli operators that act on the orbital charge states of the DQD electron, σ_x and σ_z are Pauli operators that act on the spin states of the electron, $B_z = B_z^{\text{ext}} + B_z^M$ denotes the total magnetic field along the z axis and $B_x^M = (B_{x,R}^M - B_{x,L}^M)/2$ is half the magnetic field difference of the DQD in the x direction. In the theoretical model, we have assumed that the average magnetic

field in the x direction satisfies $(B_{x,R}^M + B_{x,L}^M)/2 = 0$, which is a good approximation given the geometry of the micromagnet and its alignment with the DQD. We add the electric-dipole coupling to the cavity with the Hamiltonian

$$H_1 = g_c(a + a^\dagger)\tau_z$$

where a and a^\dagger are the photon operators for the cavity. The electric-dipole operator can be expressed in the eigenbasis $\{|n\rangle\}$ of H_0 as

$$\tau_z = \sum_{n,m=0}^3 d_{nm}|n\rangle\langle m|$$

We then write the quantum Langevin equations for the operators a and $\sigma_{nm} = |n\rangle\langle m|$:

$$\begin{aligned} \dot{a} &= i\Delta_0 a - \frac{\kappa}{2}a + \sqrt{\kappa_1}a_{\text{in},1} + \sqrt{\kappa_2}a_{\text{in},2} - ig_c e^{i\omega_R t} \sum_{n,m=0}^3 d_{nm}\sigma_{nm} \\ \dot{\sigma}_{nm} &= -i(E_m - E_n)\sigma_{nm} - \sum_{n',m'} \gamma_{nm,n'm'}\sigma_{n'm'} + \sqrt{2\gamma}F \\ &\quad - ig_c(ae^{-i\omega_R t} + a^\dagger e^{i\omega_R t})d_{nm}p_{nm} \end{aligned} \quad (2)$$

where $\Delta_0 = \omega_R - \omega_c$ is the detuning of the driving field frequency ($\omega_R = 2\pi f$) relative to the cavity frequency ($\omega_c = 2\pi f_c$) and $p_{nm} = p_n - p_m$ is the population difference between levels n and m (p_n can, for example, be assumed to be a Boltzmann distribution in thermal equilibrium). This description is equivalent to a more general master-equation approach in the weak-driving regime in which population changes in the DQD can be neglected. Furthermore, κ_1 and κ_2 are the photon decay rates at ports 1 and 2 of the cavity and $a_{\text{in},1}$ is the input field of the cavity, which we assume to couple through port 1 only ($a_{\text{in},2} = 0$). The quantum noise of the DQD F is neglected in what follows. The super-operator γ with matrix elements $\gamma_{nm,n'm'}$ represents decoherence processes, including charge relaxation and dephasing due to charge noise (these processes also imply some degree of spin relaxation and dephasing due to spin-charge hybridization via B_x^M). Our goal is to relate the incoming parts $a_{\text{in},i}$ of the external field at the ports to the outgoing fields $a_{\text{out},i} = \sqrt{\kappa_i}a - a_{\text{in},i}$. The transmission $A = \bar{a}_{\text{out},2}/\bar{a}_{\text{in},1}$ (where the overbars denote time-averaged expectation values) through the microwave cavity is then computed using a rotating-wave approximation to eliminate the explicit time dependence in equation (2), by solving the equations for the expected value of these operators in the stationary limit (\bar{a} and $\bar{\sigma}_{n,m}$):

$$A = \frac{-i\sqrt{\kappa_1\kappa_2}}{-\Delta_0 - i\kappa/2 + g_c \sum_{n=0}^2 \sum_{j=1}^{3-n} d_{n,n+j}\chi_{n,n+j}}$$

where $\chi_{n,n+j} = \bar{\sigma}_{n,n+j}/\bar{a}$ are the single-electron partial susceptibilities and d_{ij} are the dipole-transition matrix elements between DQD states.

Theoretical models for spin-photon coupling and spin decoherence. Here we present a brief derivation of the analytical expressions for the spin-photon coupling rate g_s and the spin decoherence rate γ_s . A more extensive discussion of spin-photon coupling and spin decoherence specific to our device architecture is presented in ref. 41. We focus on the $\varepsilon = 0$ regime used in Fig. 3b. Accounting for spin-charge hybridization due to the field gradient B_x^M , the relevant eigenstates of the DQD Hamiltonian in equation (1) are $|0\rangle \approx |-, \downarrow\rangle$, $|1\rangle = \cos(\Phi/2)|-, \uparrow\rangle + \sin(\Phi/2)|+, \downarrow\rangle$, $|2\rangle = \sin(\Phi/2)|-, \uparrow\rangle - \cos(\Phi/2)|+, \downarrow\rangle$ and $|3\rangle \approx |+, \uparrow\rangle$. Here we have introduced a mixing angle $\Phi = \tan^{-1}[g\mu_B B_x^M/(2t_c - g\mu_B B_z)]$. The dipole-transition matrix element for the primarily spin-like transition between $|0\rangle$ and $|1\rangle$ is $d_{01} \approx -\sin(\Phi/2)$ and the dipole-transition matrix element for the primarily charge-like transition between $|0\rangle$ and $|2\rangle$ is $d_{02} \approx \cos(\Phi/2)$. The transition between $|0\rangle$ and $|3\rangle$ is too high in energy (off-resonance) and is therefore excluded from our model. The spin-photon coupling rate is $g_s = g_c|d_{01}| = g_c|\sin(\Phi/2)|$, in agreement with previous theoretical results^{33,34}.

To calculate the effective spin decoherence rate $\gamma_s^{(\text{c})}$ that arises from charge decoherence, we first construct the operators $\sigma_{01} = |0\rangle\langle 1| \approx \cos(\Phi/2)\sigma_s + \sin(\Phi/2)\sigma_\tau$ and $\sigma_{02} = |0\rangle\langle 2| \approx \sin(\Phi/2)\sigma_s - \cos(\Phi/2)\sigma_\tau$. Here $\sigma_s = |-, \downarrow\rangle\langle -, \downarrow| + |+, \downarrow\rangle\langle +, \downarrow|$ are lowering operators for the electron spin and charge, respectively. Assuming that the electron charge states have a constant decoherence rate $\gamma_c = \gamma_1/2 + \gamma_\phi$, where γ_1 is the charge relaxation rate and γ_ϕ is a dephasing rate due to charge noise⁵⁶, the equations of motion for these operators are

$$\begin{aligned} \dot{\sigma}_{01} &= \gamma_c \left[-\sin^2\left(\frac{\Phi}{2}\right)\sigma_{01} + \frac{\sin(\Phi)}{2}\sigma_{02} \right] \\ \dot{\sigma}_{02} &= \gamma_c \left[\frac{\sin(\Phi)}{2}\sigma_{01} - \cos^2\left(\frac{\Phi}{2}\right)\sigma_{02} \right] \end{aligned}$$

Combined with charge–photon coupling, the overall equations of motion (equation (2)) in a rotating frame with a drive frequency $f \approx f_c$ assume the form

$$\begin{aligned} \dot{a} &= i\Delta_0 a - \frac{\kappa}{2}a + \sqrt{\kappa_1}a_{\text{in},1} - ig_c(d_{01}\sigma_{01} + d_{02}\sigma_{02}) \\ \dot{\sigma}_{01} &= -i\delta_1\sigma_{01} - \gamma_c \sin^2\left(\frac{\Phi}{2}\right)\sigma_{01} + \gamma_c \frac{\sin(\Phi)}{2}\sigma_{02} - ig_c ad_{10} \\ \dot{\sigma}_{02} &= -i\delta_2\sigma_{02} - \gamma_c \cos^2\left(\frac{\Phi}{2}\right)\sigma_{02} + \gamma_c \frac{\sin(\Phi)}{2}\sigma_{01} - ig_c ad_{20} \end{aligned}$$

The δ_1 and δ_2 terms are defined as $\delta_1/(2\pi) = (E_1 - E_0)/h - f$ and $\delta_2/(2\pi) = (E_2 - E_0)/h - f$, where $E_{0,1,2}$ correspond to the energy of the $|0\rangle$, $|1\rangle$ and $|2\rangle$ state, respectively. Steady-state solutions to the above equations give the electric susceptibility of the spin qubit transition $\chi_{0,1} = \sigma_{01}/a = g_s/(\delta_1 - i\gamma_s^{(c)})$, where we have identified a charge-induced spin decoherence rate $\gamma_s^{(c)} = \gamma_c[\delta_2 \sin^2(\Phi/2) + \delta_1 \cos^2(\Phi/2)]/\delta_2$. To account for spin dephasing due to fluctuations of the ^{29}Si nuclear spin bath, we express the total spin decoherence rate assuming a Voigt profile:

$$\gamma_s = \frac{\gamma_s^{(c)}}{2} + \sqrt{\left(\frac{\gamma_s^{(c)}}{2}\right)^2 + 8 \ln 2 \left(\frac{1}{T_{2,\text{nuclear}}^*}\right)^2}$$

where $T_{2,\text{nuclear}}^* \approx 1 \mu\text{s}$ is the electron-spin dephasing time due to nuclear field fluctuations^{11,38}.

When fitting the data in Fig. 3b, we use the experimentally determined values of $g_c/(2\pi) = 40 \text{ MHz}$ and $\gamma_c/(2\pi) = 35 \text{ MHz}$, along with the best-fitting field gradient $B_x^M = 15 \text{ mT}$. For every t_c , the fitted value for B_z is adjusted so that the spin-qubit frequency $(E_1 - E_0)/h$ matches the cavity frequency f_c exactly. The slight discrepancy between theory and experiment for γ_s could be due to the frequency dependence of γ_c , changes in γ_c with B_z^{ext} or other decoherence mechanisms that are not captured by this simple model. To resolve such a discrepancy, a complete measurement of γ_c as a function of $2t_c/h$ and the external field B_z^{ext} is needed.

The complete theory⁴¹ also allows $g_s/(2\pi)$ to be calculated for non-zero values of ε . Using $2t_c/h = 7.4 \text{ GHz}$, we estimate $g_s/h = 2.3 \text{ MHz}$ at $\varepsilon = 20 \mu\text{eV}$ (about 4.8 GHz), close to the value of $g_s/h = 1.0 \text{ MHz}$ measured at this DQD detuning (Fig. 3a).

In this theoretical model, we have ignored Purcell decay of the spin qubit through the cavity²⁹. This is justified because γ_s at every value of t_c is measured with a large spin–cavity detuning $\Delta \approx 10g_s$. The expected Purcell decay rate of the spin qubit is $\Gamma_p/(2\pi) = [\kappa g_s^2/(\kappa^2/4 + \Delta^2)]/(2\pi) \approx 0.02 \text{ MHz}$, well below the measured values of $\gamma_s/(2\pi)$. We also note that, at least in the $2t_c \gg E_z$ limit, spin decoherence at $\varepsilon = 0$ is dominated by noise-induced dephasing rather than by energy relaxation. This is because at $2t_c/h = 11.1 \text{ GHz}$ the spin decoherence rate $\gamma_s/(2\pi) = 0.41 \text{ MHz}$ corresponds to a coherence time of $T_2 = 0.4 \mu\text{s} \ll 2T_1 = 6.4 \mu\text{s}$. **Line shapes of vacuum Rabi splittings.** In contrast to charge–photon systems^{19,20,23}, the two resonance modes in the vacuum Rabi splittings (Fig. 2b, c) show slightly unequal widths. This effect can be seen by comparing the observed spectrum of DQD1 with the expected behaviour of an equivalent two-level charge qubit that is coupled strongly to a cavity, calculated using a master-equation simulation with thermal photon number $n_{\text{th}} = 0.02$ (black dashed line in Extended Data Fig. 7). The unequal widths are unlikely to be a result of a large thermal photon number in the cavity, because the transmission spectrum calculated with $n_{\text{th}} = 0.5$ (orange dashed line) clearly does not fit the experimental data⁵⁷.

Instead, the observed asymmetry probably arises from the dispersive interaction between the cavity and the primarily charge-like transition between $|0\rangle$ and $|2\rangle$, which results in three-level dynamics that is more complicated than the two-level dynamics that characterizes charge–photon systems. A more complete treatment of this effect is given in ref. 41. Here we compare the experimental observation with theory by calculating $A(f)/A_0$ using $g_c/(2\pi) = 40 \text{ MHz}$ (DQD1) or $g_c/(2\pi) = 37 \text{ MHz}$ (DQD2), $\gamma_c/(2\pi) = 105 \text{ MHz}$ (DQD1) or $\gamma_c/(2\pi) = 130 \text{ MHz}$ (DQD2), $\kappa/(2\pi) = 1.8 \text{ MHz}$, tunnel couplings $2t_c/h = 7.4 \text{ GHz}$, $B_x^M = 15 \text{ mT}$ and $B_z = 209.6 \text{ mT}$. The results are shown as black solid lines alongside experimental data in Extended Data Fig. 7. The agreement between experiment and theory is very good for both devices. In particular, the asymmetry between the vacuum Rabi modes is also seen in the theoretical calculations. The larger values of γ_c used in

the theoretical calculations may again be due to the frequency dependence of γ_c or to changes in γ_c with B_z^{ext} . Further experiments are needed to resolve this difference.

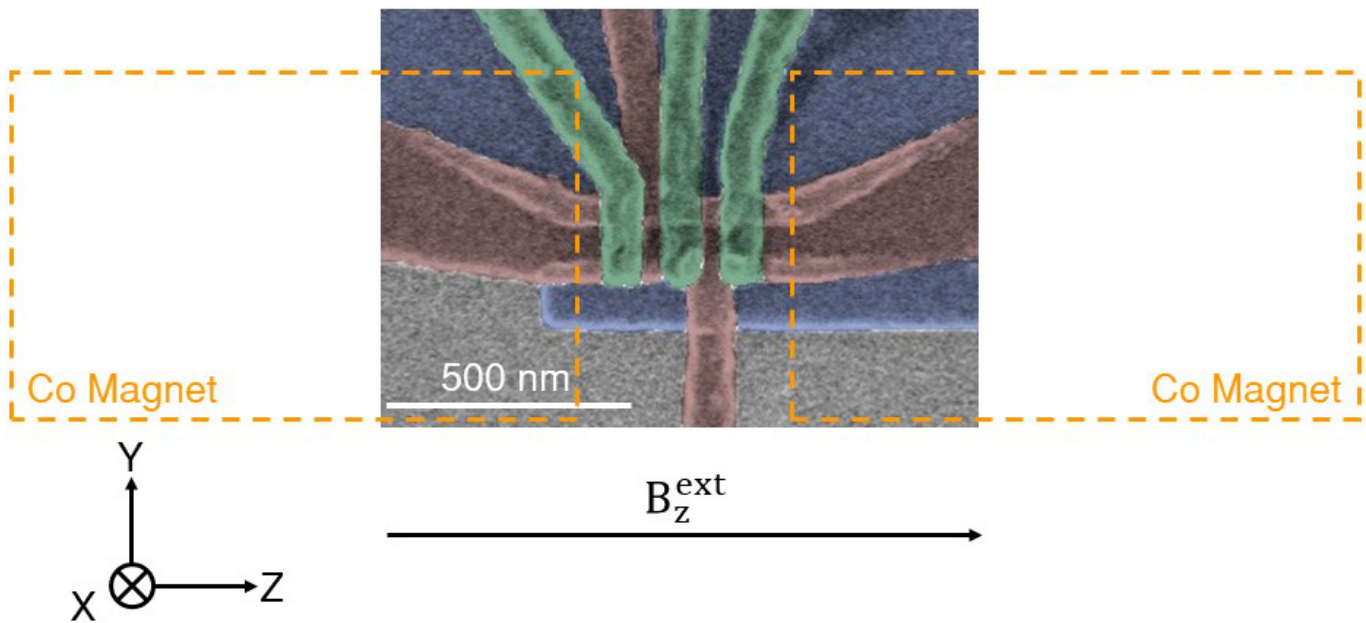
Prospects for long-range spin–spin coupling. The coherent spin–photon interface may be readily applied to enable spin–spin coupling across the cavity bus. Here we evaluate two possible schemes for implementing such a coupling, both of which have been demonstrated with superconducting qubits^{46,47}. The first approach uses direct photon exchange to perform quantum-state transfer between two qubits⁴⁷. The transfer protocol starts by tuning qubit 1 into resonance with the unpopulated cavity for a time $1/(4g_s)$, at the end of which the state of qubit 1 is transferred completely to the cavity. Qubit 1 is then detuned rapidly from the cavity and qubit 2 is brought into resonance with the cavity for a time $1/(4g_s)$, at the end of which the state of qubit 1 is transferred completely to qubit 2. Therefore, the time required for quantum-state transfer across the cavity is $1/(2g_s)$. Because the decay of vacuum Rabi oscillations occurs at a rate $\kappa/2 + \gamma_s$, the threshold for coherent-state transfer between two spin qubits is $2g_s/(\kappa/2 + \gamma_s) > 1$. The ratio $2g_s/(\kappa/2 + \gamma_s)$ is plotted as a function of $2t_c/h$ in Extended Data Fig. 8a. It can be seen that $2g_s/(\kappa/2 + \gamma_s) > 1$ for all values of $2t_c/h$, indicating that spin–spin coupling via real photon exchange is achievable and may be implemented at any value of t_c . For our device parameters, the regime $2t_c/h \approx 6 \text{ GHz}$, in which spin–charge hybridization is large and the ratio $2g_s/(\kappa/2 + \gamma_s)$ reaches a maximum of 3.5, seems most advantageous for such a coupling scheme.

The second approach to spin–spin coupling uses virtual photon exchange⁴⁶. In this scheme, both spin qubits would operate in the dispersive regime, with an effective coupling rate of $J = g_s^2(1/\Delta_1 + 1/\Delta_2)/2$, where Δ_1 and Δ_2 are the qubit–cavity detunings for qubits 1 and 2, respectively. Assuming that both qubits operate with an equal detuning $\Delta_{1,2} = 10g_s$ to minimize Purcell decay, $J = g_s/10$. For coherent spin–spin interaction, $J > \gamma_s$ needs to be satisfied, leading to the condition $g_s/\gamma_s > 10$. In Extended Data Fig. 8b, we plot the ratio g_s/γ_s as a function of $2t_c/h$, observing a maximum of $g_s/\gamma_s \approx 4$ at $2t_c/h \approx 10 \text{ GHz}$. Because the dominant spin mechanism is probably hyperfine-induced dephasing by the ^{29}Si nuclei in this regime (the decoherence rate $\gamma_s/(2\pi) \approx 0.4 \text{ MHz}$ is close to the decoherence rates commonly found with single-spin qubits in natural Si; ref. 38), transitioning to isotopically purified ^{28}Si host materials is likely to lead to an order-of-magnitude reduction in $\gamma_s/(2\pi)$, as demonstrated recently⁵⁸. Such an improvement will allow virtual-photon-mediated spin–spin coupling to be implemented in our device architecture as well.

Last, we note that both coupling approaches will benefit substantially from larger values of the charge–photon coupling rate g_c , which is achievable through the development of higher-impedance cavities^{20,59}. The superconducting cavity used here is estimated to have an impedance between 200Ω and 300Ω . Increasing this value to about $2 \text{ k}\Omega$, which is possible by using NbTiN as the superconducting material, will lead to another factor-of-three increase in g_c and therefore g_s . This could enable the $g_s/\gamma_s > 100$ regime to be accessed, where high-fidelity two-qubit gates can be implemented between distant spins. Improvements in the fidelity of cavity-mediated two-qubit gates, particularly in the case of real photon exchange, can also be sought by improving the quality factor of the cavity (and thereby reducing κ). This is achievable by implementing stronger gate line filters⁴³ and removing lossy dielectrics such as the atomic-layer-deposited Al_2O_3 underneath the cavity.

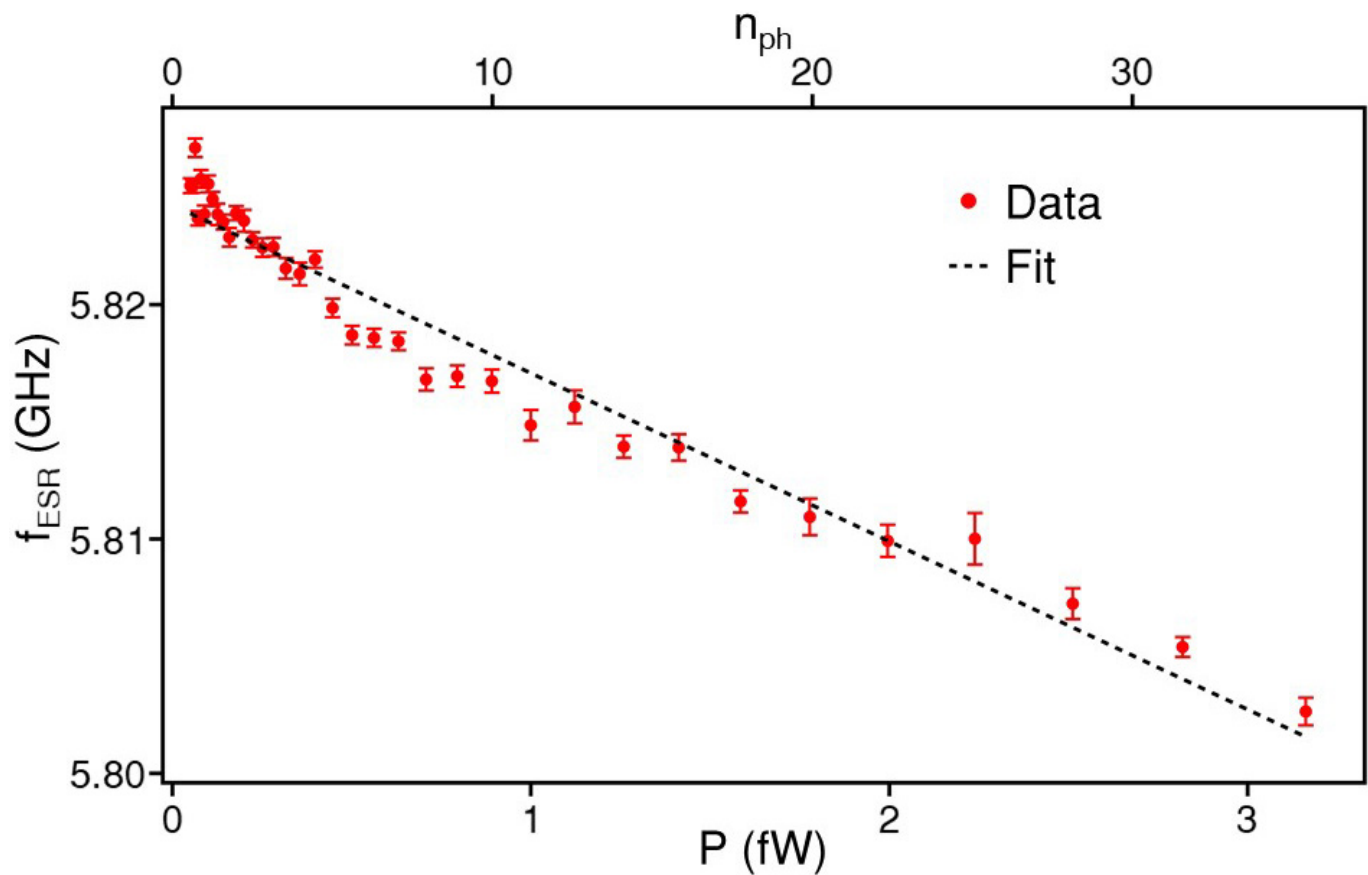
Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request. Source Data for Figs 1–4 and Extended Data Figs 2–8 are available with the online version of the paper.

55. Wallraff, A. *et al.* Approaching unit visibility for control of a superconducting qubit with dispersive readout. *Phys. Rev. Lett.* **95**, 060501 (2005).
56. Wallraff, A., Stockklauser, A., Ihn, T., Petta, J. R. & Blais, A. Comment on “Vacuum Rabi splitting in a semiconductor circuit QED system”. *Phys. Rev. Lett.* **111**, 249701 (2013).
57. Rau, I., Johansson, G. & Shnirman, A. Cavity quantum electrodynamics in superconducting circuits: Susceptibility at elevated temperatures. *Phys. Rev. B* **70**, 054521 (2004).
58. Yoneda, J. *et al.* A >99.9% fidelity quantum-dot spin qubit with coherence limited by charge noise. Preprint at <https://arxiv.org/abs/1708.01454> (2017).
59. Samkharadze, N. *et al.* High-kinetic-inductance superconducting nanowire resonators for circuit QED in a magnetic field. *Phys. Rev. Appl.* **5**, 044004 (2016).



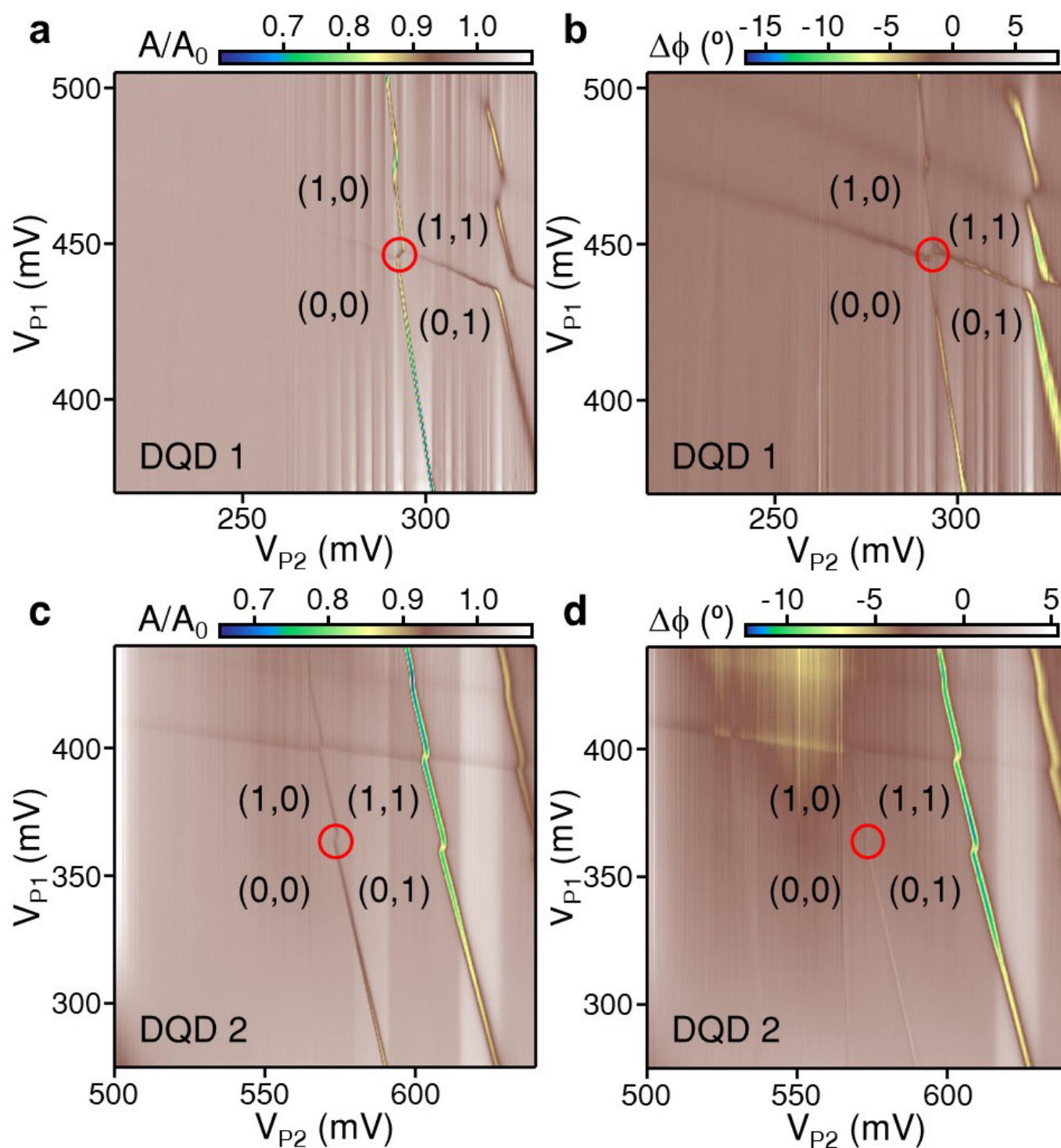
Extended Data Figure 1 | Micromagnet design. To-scale drawing of the micromagnet design, superimposed on top of the SEM image of the DQD. The coordinate axes and the direction of the externally applied magnetic field B_z^{ext} are indicated at the bottom. In this geometry, the DQD electron experiences a homogeneous z field $B_z \approx B_z^{\text{ext}} + B_z^{\text{M}}$. The total x field B_x that

is experienced by the electron is spatially dependent, being approximately $B_{x,L}^{\text{M}}$ ($B_{x,R}^{\text{M}}$) when the electron is in the L (R) dot ($|\varepsilon| \gg t_c$) and $(B_{x,L}^{\text{M}} + B_{x,R}^{\text{M}})/2$ when the electron is delocalized between the DQDs ($\varepsilon = 0$). The y field B_y for the DQD electron is expected to be small compared to the other field components for this magnet design.



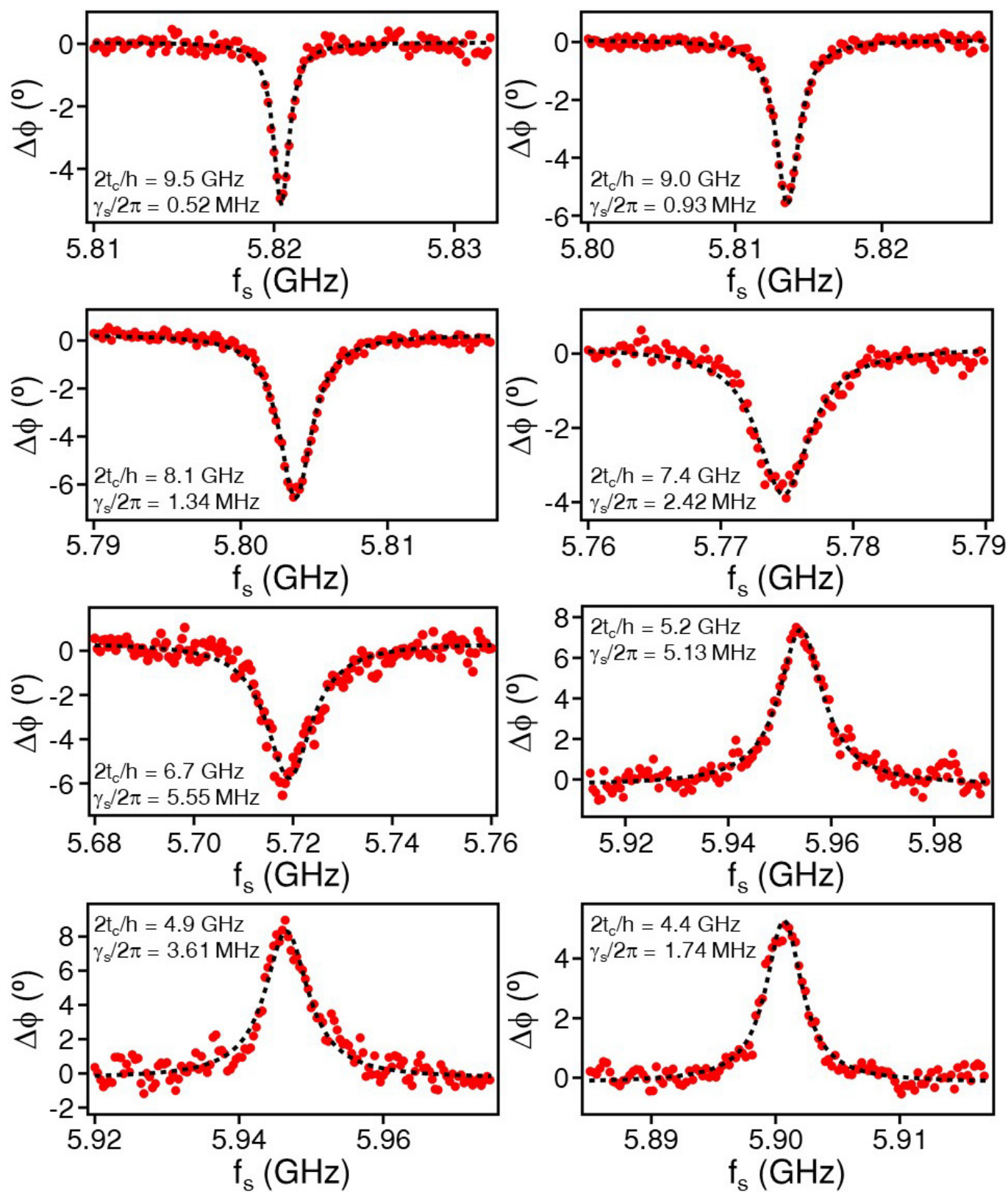
Extended Data Figure 2 | Photon number calibration. The ESR resonance frequency f_{ESR} , measured using the phase response of the cavity $\Delta\phi$ in the dispersive regime (Fig. 4b), is plotted as a function of the estimated power at the input port of the cavity P (data). The device is configured with $g_s/(2\pi) = 2.4$ MHz and spin-photon detuning $\Delta/(2\pi) \approx -18$ MHz. The dashed line shows a fit to

$f_{\text{ESR}} = f_{\text{ESR}}(P=0) + (2n_{\text{ph}}g_s^2/\Delta)/(2\pi)$, where n_{ph} is the average number of photons in the cavity, plotted as the top x axis. The experiments are conducted with $P \approx -133$ dBm (0.05 fW), which corresponds to $n_{\text{ph}} \approx 0.6$. The error bars indicate the uncertainties in the centre frequency of the ESR transition.



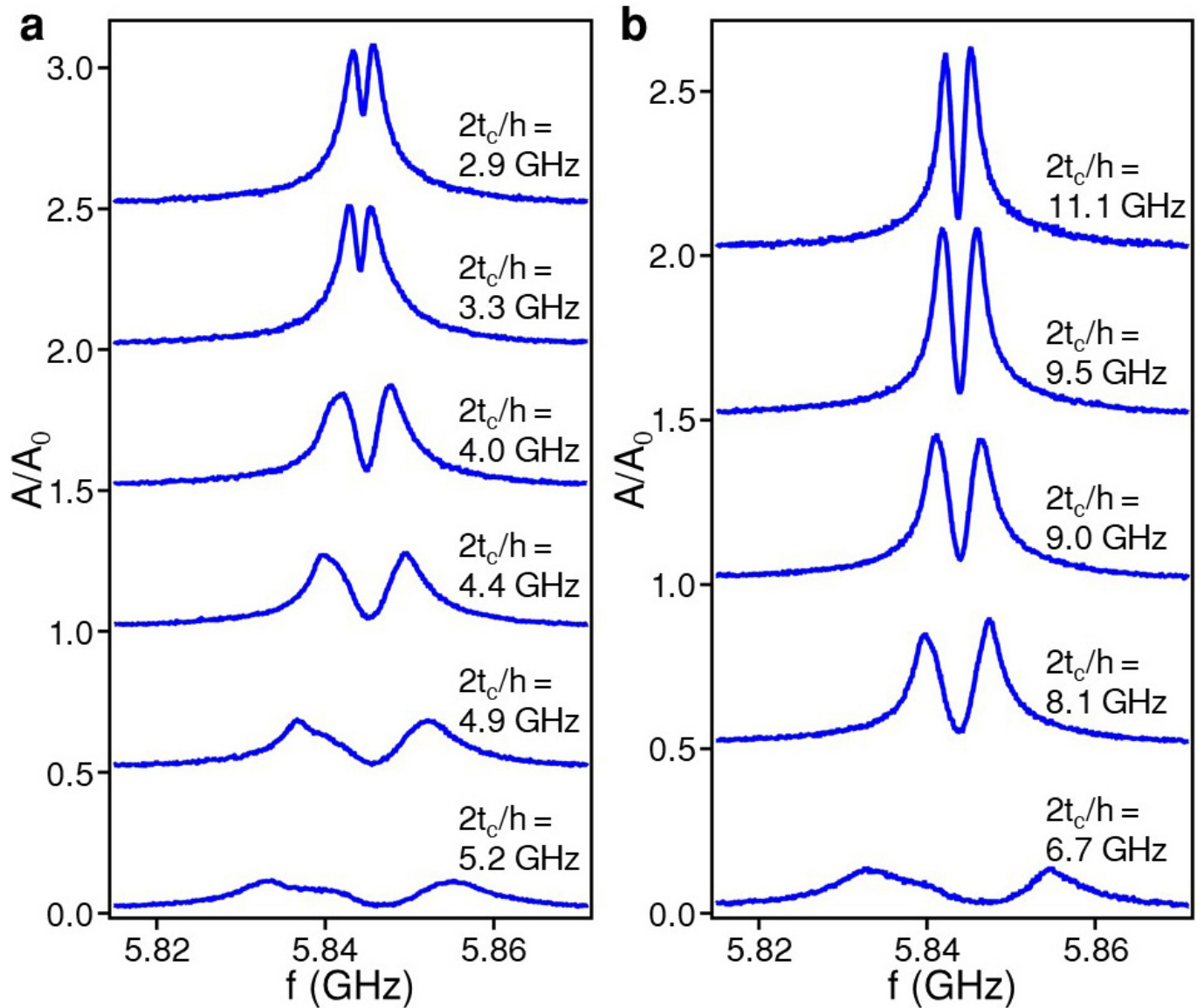
Extended Data Figure 3 | DQD stability diagrams. The cavity transmission amplitude A/A_0 (a, c) and phase response $\Delta\phi$ (b, d) are plotted as functions of V_{P1} and V_{P2} for DQD1 (a, b) and DQD2 (c, d), obtained with $f=f_c$. The (1, 0) \leftrightarrow (0, 1) transitions are clearly identified

on the basis of these measurements and subsequently tuned close to resonance with the cavity for the experiments described in the main text. The red circles indicate the locations of the (1, 0) \leftrightarrow (0, 1) transitions of the two DQDs.



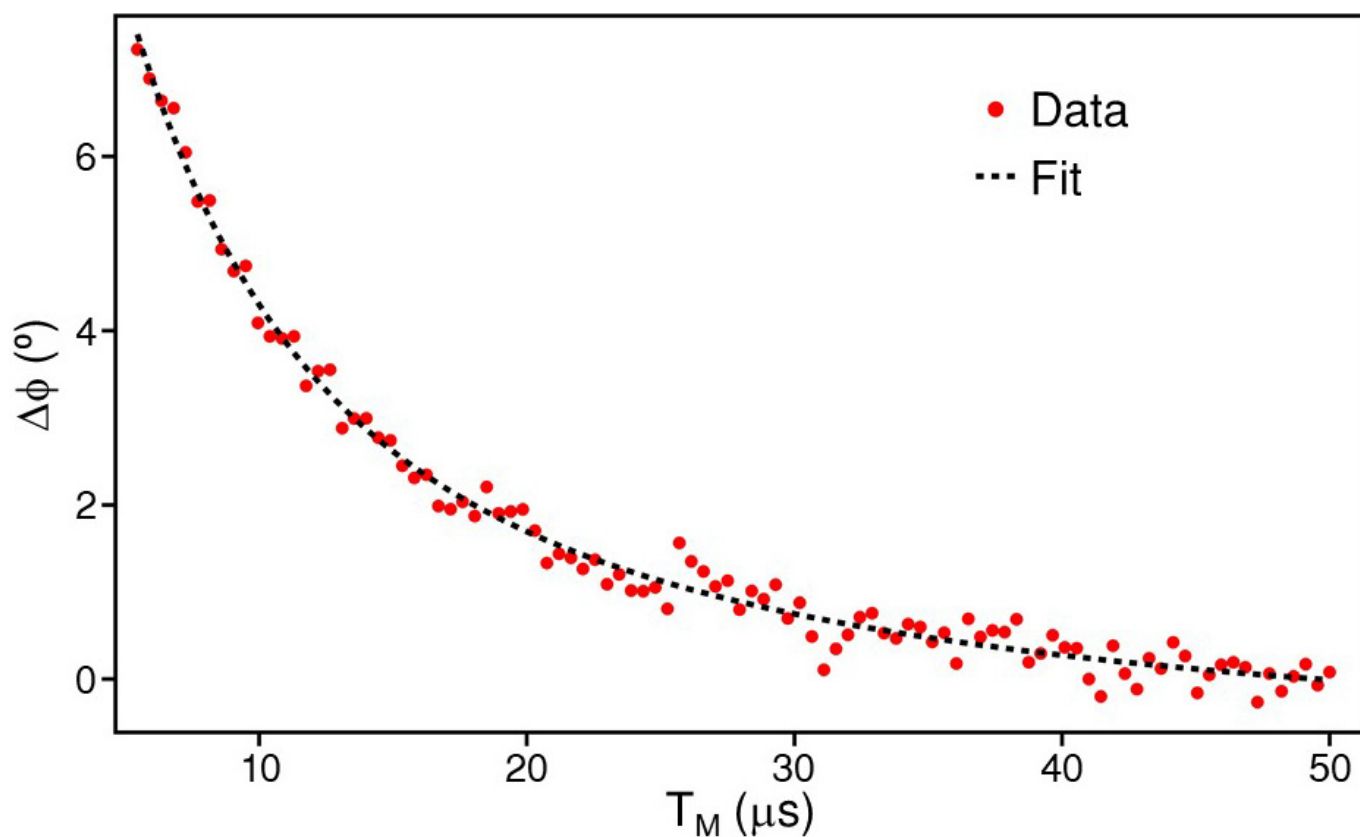
Extended Data Figure 4 | Spin decoherence rates at different DQD tunnel couplings. ESR line, as measured in the cavity phase response $\Delta\phi(f_s)$, is shown for different values of $2t_c/h$ in the low-power limit (data). $\varepsilon = 0$ for every dataset. Dashed lines are fits with Lorentzian functions

and $\gamma_s/(2\pi)$ is determined as the half-width at half-maximum of each Lorentzian. The spin-photon detuning $|\Delta| \approx 10 g_s$ for each dataset, to ensure that the system is in the dispersive regime.



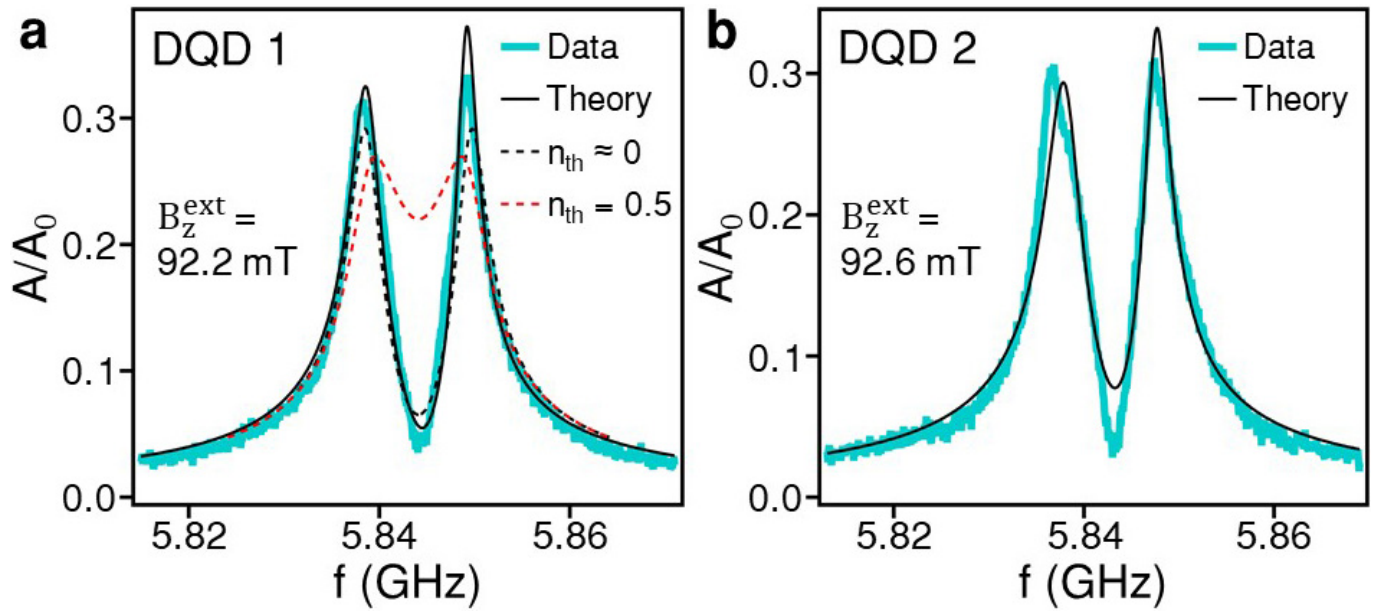
Extended Data Figure 5 | Spin-photon coupling strengths at different DQD tunnel couplings. a, b, Vacuum Rabi splittings for $2t_c/h < f_c$ (a) and $2t_c/h > f_c$ (b), obtained by varying B_z^{ext} until a pair of resonance peaks with approximately equal heights emerges in the cavity transmission

spectrum $A/A_0 \cdot g_s/(2\pi)$ is then estimated as half the frequency difference between the two peaks. $\varepsilon = 0$ for every dataset. g_s is difficult to measure for $5.2 \text{ GHz} < 2t_c/h < 6.7 \text{ GHz}$ owing to the small values of A/A_0 that arise from the large spin decoherence rates γ_s in this regime.



Extended Data Figure 6 | Spin relaxation at $\varepsilon = 0$. The time-averaged phase response of the cavity $\Delta\phi$ is shown as a function of wait time T_M (data), measured using the pulse sequence illustrated in Fig. 4c. The microwave burst time is fixed at $\tau_B = 80$ ns. The dashed line shows a fit

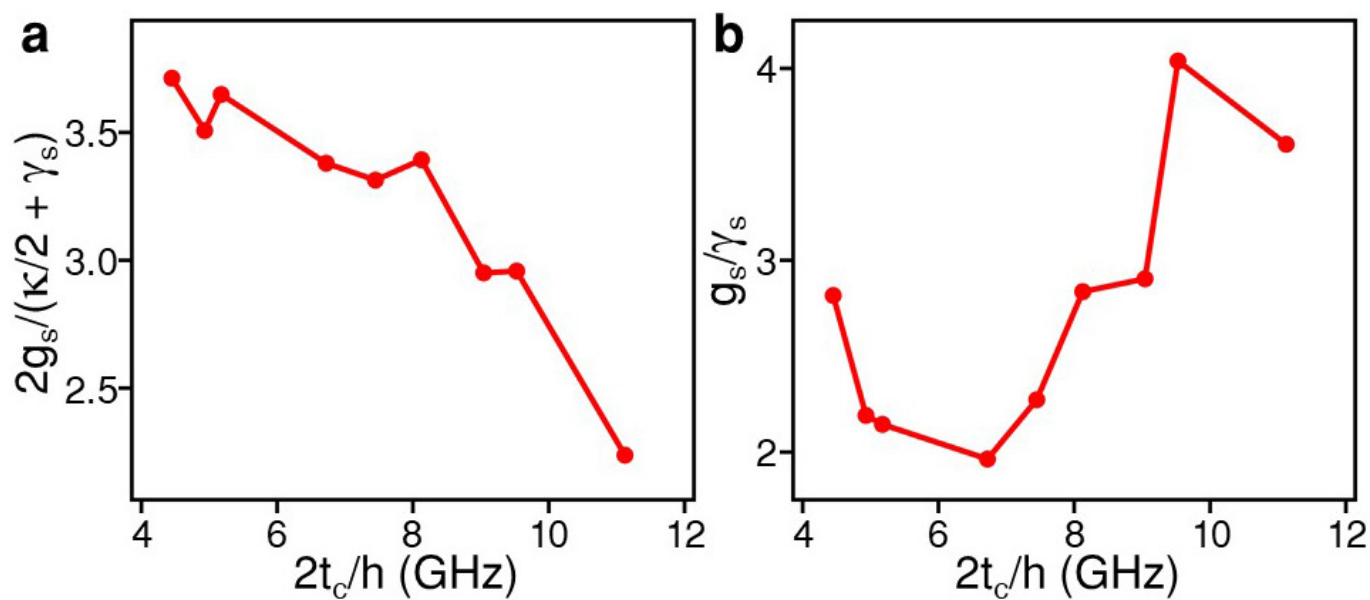
using the function $\phi_0 + \phi_1(T_1/T_M)[1 - \exp(-T_M/T_1)]$, which yields a spin relaxation time of $T_1 \approx 3.2 \mu\text{s}$. The experimental conditions are the same as for Fig. 4d.



Extended Data Figure 7 | Theoretical fits to vacuum Rabi splittings.

The calculated cavity transmission spectra (black solid lines) are superimposed on the experimentally measured vacuum Rabi splittings shown in Fig. 2b, c (data). The calculations are produced with $g_c/(2\pi) = 40$ MHz ($g_c/(2\pi) = 37$ MHz), $\kappa/(2\pi) = 1.8$ MHz, $\gamma_c/(2\pi) = 105$ MHz ($\gamma_c/(2\pi) = 120$ MHz), $B_z = B_z^{\text{ext}} + B_z^{\text{M}} = 209$ mT,

$B_x^{\text{M}} = (B_{x,\text{R}}^{\text{M}} - B_{x,\text{L}}^{\text{M}})/2 = 15$ mT and $2t_c/h = 7.4$ GHz for DQD1 (DQD2). For comparison, $A(f)/A_0$, simulated for a two-level charge qubit with a decoherence rate of $\gamma_c/(2\pi) = 2.4$ MHz coupled to a cavity with $\kappa/(2\pi) = 1.8$ MHz at a rate $g_c/(2\pi) = 5.5$ MHz, is shown in a for thermal photon numbers of $n_{\text{th}} = 0.02$ (black dashed line) and $n_{\text{th}} = 0.5$ (red dashed line).



Extended Data Figure 8 | Prospect for long-range spin-spin coupling. **a**, The ratio $2g_s/(\kappa/2 + \gamma_s)$ as a function of $2t_c/h$, calculated using the data in Fig. 3b and $\kappa/(2\pi) = 1.8$ MHz. **b**, The ratio g_s/γ_s as a function of $2t_c/h$, also calculated using the data in Fig. 3b.

Planning chemical syntheses with deep neural networks and symbolic AI

Marwin H. S. Segler^{1,2}, Mike Preuss³ & Mark P. Waller⁴

To plan the syntheses of small organic molecules, chemists use retrosynthesis, a problem-solving technique in which target molecules are recursively transformed into increasingly simpler precursors. Computer-aided retrosynthesis would be a valuable tool but at present it is slow and provides results of unsatisfactory quality. Here we use Monte Carlo tree search and symbolic artificial intelligence (AI) to discover retrosynthetic routes. We combined Monte Carlo tree search with an expansion policy network that guides the search, and a filter network to pre-select the most promising retrosynthetic steps. These deep neural networks were trained on essentially all reactions ever published in organic chemistry. Our system solves for almost twice as many molecules, thirty times faster than the traditional computer-aided search method, which is based on extracted rules and hand-designed heuristics. In a double-blind AB test, chemists on average considered our computer-generated routes to be equivalent to reported literature routes.

Retrosynthetic analysis is the canonical technique used to plan the synthesis of small organic molecules^{1,2}. In retrosynthesis, a search tree is built by ‘working backwards’, analysing molecules recursively and transforming them into simpler precursors until one obtains a set of known or commercially available building-block molecules (Fig. 1)^{3,4}. Given that transformations are formally reversed chemical reactions, the plan can be then carried out in the laboratory in the forward direction to synthesize the target compound^{3,4}. Transformations are derived from successfully conducted series of similar reactions with analogous starting materials, and are often named after their discoverers (‘named reactions’)⁵. At each retrosynthetic step, a small set out of hundreds of thousands of transformations known in modern chemistry has to be selected. In a pattern recognition process, chemists intuitively prioritise the most promising transformations, which they then consider, without actively thinking about the less promising ones⁶. However, when a transformation is applied to a new molecule, there is no guarantee that the corresponding reaction will proceed in the expected way⁷. A molecule failing to react as predicted is called ‘out of scope’. This can be due to steric or electronic effects, an incomplete understanding of the reaction mechanism, or conflicting reactivity in the molecular context. Predicting which molecules are ‘in scope’ can be challenging even for the best human chemists^{4,7}.

Computer-assisted synthesis planning (CASP) could help chemists to find better routes faster, and is a missing component in virtual *de novo* design and robot systems performing molecular design–synthesis–test cycles^{8–10}. To perform CASP, the knowledge that humans gain must be transferred into an executable program^{11–16}. Despite 60 years of research, attempts to formalize chemistry by manual encoding by experts have not convinced synthetic chemists, and it does not scale to exponentially growing knowledge^{15–19}. Methods of algorithmically extracting transformations from reaction datasets^{20–22} have been criticized for high noise and lack of ‘chemical intelligence’^{13,14}. However, we recently showed that deep neural networks can learn to rank extracted symbolic transformations, and to avoid reactivity conflicts, which mimics the expert’s intuitive decision-making²³. To guide the search in promising directions, heuristic best first search (BFS) has been employed, in which hand-designed heuristic functions determine

position values¹³. Unfortunately, unlike in chess, it is difficult to define strong heuristics in chemistry for three reasons. First, chemists tend to disagree on what constitutes a good position^{24,25}. Second, although it is generally desirable to simplify the molecules, it can be tactically beneficial to temporarily increase complexity by the use of protecting or directing groups. Finally, the position value depends highly on the availability of suitable precursors^{13,15}. Even complex molecules can be made in a few steps if precursors are readily available. Therefore, one cannot reliably estimate the value of a synthetic position without completely ‘playing the molecules until the end of the game’.

Monte Carlo tree search (MCTS) has emerged as a general search technique for sequential decision problems with large branching factors without strong heuristics, such as games or automated theorem proving^{26–28}. MCTS uses rollouts to determine position values. Rollouts are Monte Carlo simulations, in which random search steps are performed without branching until a solution has been found or a maximum depth is reached. These random steps can be sampled from machine-learned policies $p(t|s)$ ²⁹, which predict the probability of taking the move (applying the transformation) t in position s , and are trained to predict the winning move by using human games or self-play^{30–35}.

In this work, we combine three different neural networks together with MCTS to perform chemical synthesis planning (3N-MCTS). The first neural network (the expansion policy) guides the search in promising directions by proposing a restricted number of automatically extracted transformations. A second neural network then predicts whether the proposed reactions are actually feasible (in scope). Finally, to estimate the position value, transformations are sampled from a third neural network during the rollout phase. The neural networks were trained on essentially all reactions published in the history of organic chemistry.

Training the expansion and rollout policies

We extracted transformation rules from 12.4 million single-step reactions from the Reaxys³⁶ chemistry database²³. Two sets of rules were extracted. The rollout set comprises rules that contain the atoms and bonds that changed in the course of the reaction (the reaction centre),

¹Institute of Organic Chemistry and Center for Multiscale Theory and Computation, Westfälische Wilhelms-Universität, Münster, Germany. ²BenevolentAI, London, UK. ³European Research Center for Information Systems, Westfälische Wilhelms-Universität Münster, Germany. ⁴Department of Physics and International Centre for Quantum and Molecular Structures, Shanghai University, Shanghai, China.

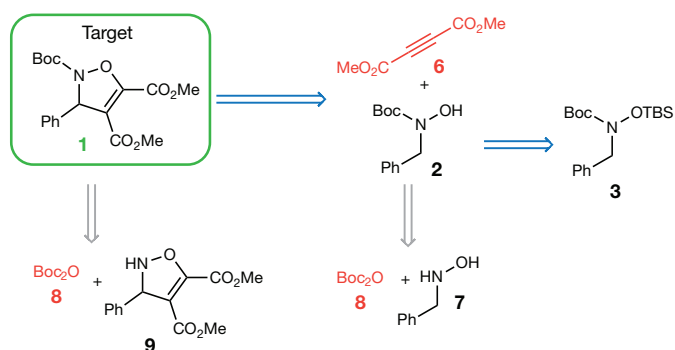
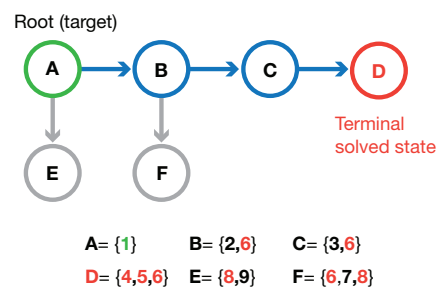
a Chemical representation of the synthesis plan**b** Search tree representation

Figure 1 | Translation of the traditional chemists' retrosynthetic route representation to the search tree representation. **a**, The traditional chemists' retrosynthetic route representation (conditions omitted)⁵⁰. **b**, The search tree representation. The nodes in the tree represent the synthetic position, and contain all precursors needed to make the molecules of the preceding positions all the way down to the tree's

root, which contains the target. Branches in the search tree correspond to complete routes. Calculating the value of branches through task-dependent scoring functions allows us to compare and rank different routes. The target molecule can be solved if it can be deconstructed to a set of readily available building blocks (marked red). Ph, phenyl; Boc, *tert*-butyloxycarbonyl; TBS, *tert*-butyldimethylsilyl.

and the first-degree neighbouring atoms. Only rules that occurred at least 50 times in reactions published before 2015 were kept. For the expansion rules, a more general rule definition was employed. Here, only the reaction centre was extracted. Rules occurring at least three times were kept. The two sets encompass 17,134 and 301,671 rules, and cover 52% and 79% of all chemical reactions from 2015 and after, respectively.

Rule extraction associates each reaction, and thus each product, with a transformation rule. This allows us to train neural networks as policies to predict the best transformations given the product, or in other words, the best reactions with which to make the product²³. Importantly, such neural networks also learn about the context in which the reactions can occur (functional group tolerance)²³. For the expansion policy, we employed a deep highway network³⁷ with exponential linear unit nonlinearities³⁸. To assess its ability to generalize, we performed a time-split strategy³⁹. For training, only reactions published before 2015 were used, whereas for validation and testing, data from 2015 and later were selected.

Extended Data Table 1 shows the metrics for the expansion policy. The neural network predicts the correct solution out of 301,671 transformations with an accuracy of 31%, which is reasonable. It has to be noted that there are almost always many feasible ways to make a molecule. The top 10 and top 50 accuracies of 63.3% and 72.5% indicate that the correct transformations are generally ranked highly. Beyond the top 50 predicted results, the accuracy increases only marginally. This observation allows us to reduce the branching factor drastically, which is 46,175 when rules are applied exhaustively. During search tree expansion, we restrict the possible transformations to a maximum of 50. Additionally, we sum the probabilities of the predicted actions, starting from the highest-ranked transformation. When the cumulative probability reaches 0.995, we stop further expansion, even if fewer than 50 actions have been expanded. This allows the system to focus on highly likely transformations when only a few good options exist, for example in the synthesis of acyl chlorides or Grignard reagents. We observed that the reactions in this reduced top 50 are almost always reasonable and are often variations of the correct prediction. For example, a Heck reaction can often be conducted with bromide, iodide or triflate as the leaving group.

The rollout policy network, which is a neural network with one hidden layer, is trained in the same way as the expansion policy. It uses a set of 17,134 rules, which implies a lower coverage than the expansion policy, yet it needs just 10 ms to make a prediction, in contrast to 90 ms for the expansion policy, owing to the smaller output layer (see Extended Data Table 1). The rationale for using two different rule sets is to use a powerful but slow policy to select the best candidate transformations for expansion, and a fast rollout policy to estimate the position values³⁵.

Prediction with the in-scope filter network

After the search space has been narrowed down by the expansion policy to the most promising transformations, we need to predict whether the corresponding reactions will actually work for a particular molecule. We trained a deep neural network as a binary classifier to predict whether the reactions corresponding to the transformations selected by the policy network are actually feasible⁴⁰. The classifier has to be trained on successful and failed reactions. Unfortunately, failed reactions are rarely reported and not contained in reaction databases. However, published reactions contain implicit information about reactions that do not occur. For a high-yielding reaction $A + B \rightarrow C$, we can assume that hypothetical products D, E, \dots are not formed. By applying reaction rules in the forward direction to the reactants of reported reactions, negative reactions, for example, with incorrect regio- and chemoselectivity, can then be generated^{41,42}. Here we used the same rule set as for the expansion policy. Additionally, we generated negative examples by shuffling the associated pairs of products and corresponding reactions (see Methods for details). Using these data augmentation strategies, we generated 100 million negative reactions from reactions published before 2015 for training and 10 million published in and after 2015 for testing. As positive cases, all reported reactions from these periods were used. On the test set, the classifier achieves an area under the receiver operation characteristic curve of 0.99, and an area under the precision-recall curve of 0.94, which indicates good performance (see Extended Data Fig. 1)⁴³. The false positive rate of the filter (that is, incorrect reactions passing the filter) is 1.5%, whereas the false negative rate (that is, real reactions being filtered out) is 14%. Interestingly, the in-scope filter correlates with basic electronic properties (Hammond parameters and lowest unoccupied molecular orbital (LUMO) energies), even though it is not explicitly trained to do so (see Methods and Extended Data Fig. 5).

Integrating neural networks and MCTS

The expansion policy network and the in-scope filter network are combined into a pipeline (Fig. 2b). When a position s_i is to be analysed, each molecule of the position is fed into the policy network. Then, the transformations with the highest scores are applied to the molecule, which yields the possible precursors and thus full reactions. These reactions are submitted to the in-scope filter, where only transformations and precursors corresponding to positively classified reactions are kept. They represent the 'legal moves' available in position s_i .

The expansion procedure and the rollout policy are then incorporated in the respective phases of an MCTS algorithm to form 3N-MCTS. The four MCTS phases are then iterated to build the search tree.

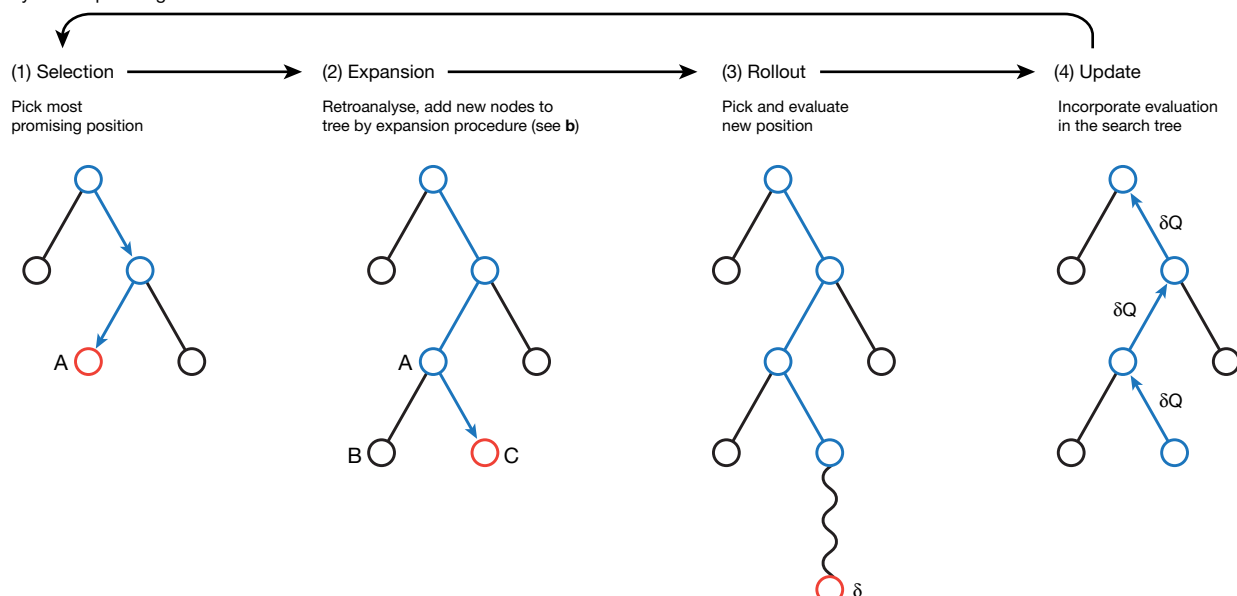
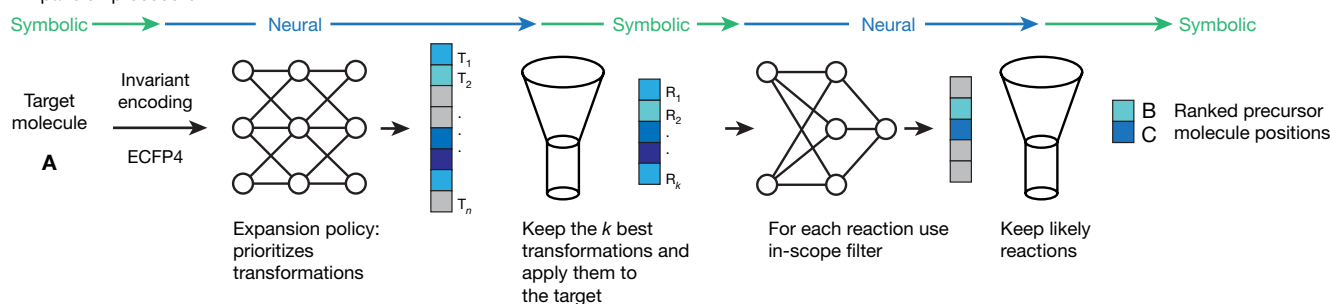
a Synthesis planning with Monte Carlo tree search**b** Expansion procedure

Figure 2 | Schematic of MCTS methodology. **a**, MCTS searches by iterating over four phases. In the selection phase (1), the most urgent node for analysis is chosen on the basis of the current position values. In phase (2) this node may be expanded by processing the molecules of the position A with the expansion procedure (**b**), which leads to new positions B and C, which are added to the tree. Then, the most promising new position is chosen, and a rollout phase (3) is performed by randomly sampling transformations from the rollout policy until all molecules are solved or a certain depth is exceeded. In the update phase (4), the position values are updated in the current branch to reflect the result of the

rollout. **b**, Expansion procedure. First, the molecule (A) to retroanalyse is converted to a fingerprint and fed into the policy network, which returns a probability distribution over all possible transformations (T_1 to T_n). Then, only the k most probable transformations are applied to molecule A. This yields the reactants necessary to make A, and thus complete reactions R_1 to R_k . For each reaction, the reaction prediction is performed using the in-scope filter, returning a probability score. Improbable reactions are then filtered out, which leads to the list of admissible actions and corresponding precursor positions B and C.

(1) Selection. In the first 3N-MCTS phase, starting at the root node (the target molecule) of the search tree, the algorithm sequentially selects the most promising next position within the tree until a leaf node is reached (Fig. 2a). The algorithm balances the selection of high-value positions and unexplored positions. If a leaf node is visited for the first time, it is directly evaluated by a rollout. If it is visited for the second time, it is expanded by processing via the expansion policy.

(2) Expansion. Now, the possible transformations determining the follow-up positions of the current position are selected by applying the expansion procedure. The predicted follow-up positions are added to the tree as children of the leaf node, and the most promising position is selected for rollout.

(3) Rollout. This phase starts with checking the status of the position. If it is already solved, the algorithm directly receives a reward greater than 1 to encourage exploitation. Non-terminal states are subjected to a rollout, where actions are sampled from the rollout network recursively, until the state has been deconstructed into building blocks or a maximal depth is reached.

(4) Update. If a solution has been found during rollout, a reward of 1 is received. Partial rewards are given if some, but not all, molecules in the state are solved. If no solution was found, a reward of -1 is received. Here, bespoke scoring functions for the problem at hand,

such as process chemistry or small-scale medicinal chemistry, can also be supplied. Eventually, the tree is updated to incorporate the achieved reward by updating the position values.

These four phases of 3N-MCTS are iterated until a time budget or maximal iteration count is exceeded. Finally, to obtain the synthesis plan, we repeatedly select the retrosynthetic step with the highest value until a solved position is reached, or a maximum depth has been exceeded, in which case the problem is unsolved.

Evaluating the performance characteristics of 3N-MCTS

To evaluate the performance of 3N-MCTS, we compare our algorithm to the state-of-the-art search method, which is BFS with the hand-coded SMILES^{3/2} heuristic cost function ('heuristic BFS')¹³. This function assigns the lowest cost to steps that split up the molecule into equally sized parts. Additionally, we perform BFS with the cost calculated by the policy network ('neural BFS'). All algorithms use the same set of automatically extracted transformations. The evaluation is again time-split, as follows. Models were trained only on data published before 2015. As test data, only molecules first reported in or after 2015 were considered (which were not contained in the training dataset). Provided with the target molecules, the algorithms then had to find a synthesis route to given building blocks.

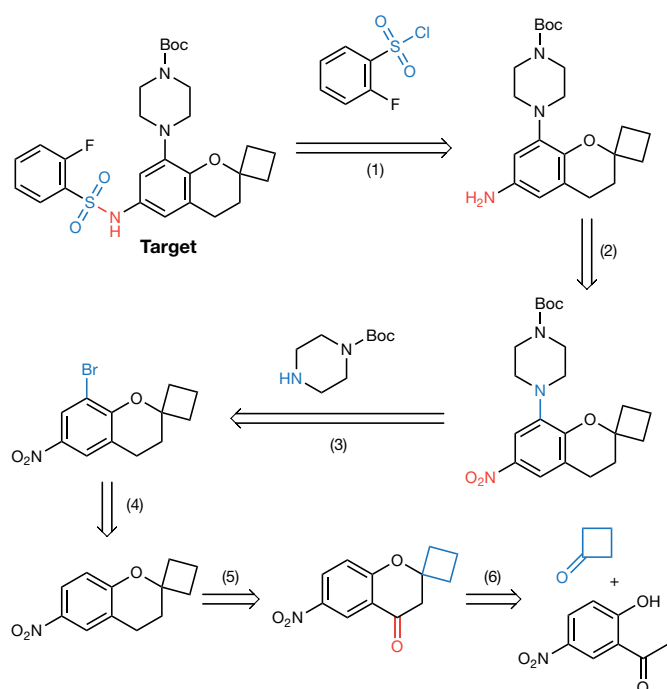


Figure 3 | An exemplary six-step synthesis route for an intermediate in a drug candidate synthesis. This route is identical to the published one⁴⁴ and was found by our algorithm autonomously within 5.4 s. The affected functional groups in each step are marked blue or red.

Figure 3 shows an exemplary six-step route for an intermediate of a drug candidate synthesis reported in 2015, which was found by our algorithm in 5.4 s. It matches the published route⁴⁴. Several hundred additional exemplary retrosynthetic routes found by the MCTS algorithm for molecules first synthesized in or after 2015 are deposited in Supplementary Information (see also Extended Data Fig. 2).

Quantitative evaluation

Surprisingly, in the past, neither hand-coded nor automatically extracted retrosynthetic systems have been validated at scale in a statistical way. We quantitatively assessed the performance characteristics of the different search algorithms by finding synthesis routes for 497 diverse molecules first reported in or after 2015 to known building blocks (see Fig. 4).

MCTS already solves more than 80% of the test set with a time limit of 5 s per target molecule, compared to 40% with neural BFS and 0% for heuristic BFS. MCTS solved 92% of the test set with a limit of 60 s per molecule, whereas neural BFS solved 71%, and heuristic BFS solved 4%. Even at much longer runtimes of 20 min per molecule, heuristic and neural BFS are not able to compete with MCTS. Provided with infinite runtime, however, the algorithms will converge to the same performance. The molecules that MCTS failed to solve could not be solved by the BFS algorithms either. When looking beyond the first (top 1) retrieved route, MCTS and BFS find similar alternative routes, and do not differ much in terms of route diversity (see Supplementary Information section 2).

To determine which MCTS components are responsible for its superior performance, we compared MCTS against several related search algorithms (see Table 1) at a runtime limit of 3×300 s (three restarts). MCTS in conjunction with the expansion policy network solved the highest number of retrosynthetic targets. On average, MCTS required the least amount of time per molecule to find a solution (entry 1). Plain Monte Carlo search randomly selects transformations using the expansion policy network, without building a tree. The Monte Carlo search (entry 2) solved 89.54% of the test set. UCT is an MCTS variant that uses the expansion policy network only to narrow down the possible transformations, but not to guide the

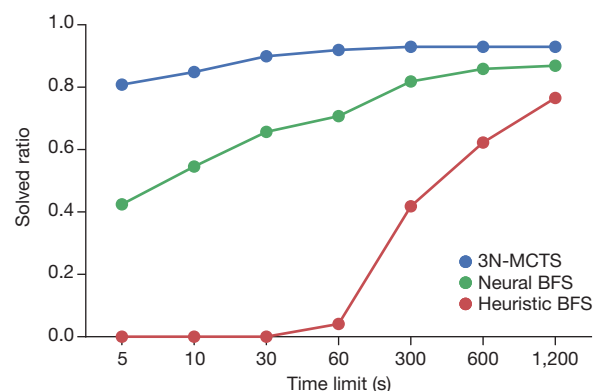


Figure 4 | Influence of the time per query on performance. The maximum number of steps is 100,000.

search via the predicted probability of the transformation²⁸. In this way, 87.12% of the test set is solved. BFS using a cost function based on the expansion policy network solved only 84.24%, highlighting the importance of rollouts. The traditional approach, BFS with a hand-designed heuristic cost function, solves only 45.6% of the test set, and needs 433.4 s on average to find a solution. These results suggest that all components of 3N-MCTS (building a tree, reducing the branching factor via the expansion policy, guiding the search with the expansion policy, and using rollouts) contribute to its superior performance. We also found 3N-MCTS to be robust towards the choice of the MCTS parameters (see Supplementary Information section 1).

Assessing route quality via double-blind AB tests

The central criticism of retrosynthesis systems has been that the proposed routes often contain what chemists immediately recognize as chemically unreasonable steps. Therefore, to assess the quality of the solutions we conducted two AB tests, in which 45 graduate-level organic chemists from two world-leading organic chemistry institutes in China and Germany had to choose one of two routes leading to the same molecule on the basis of personal preference and synthetic plausibility. The tests were double-blind, meaning that neither the participants nor the conductors were aware of the origin of the routes. The test molecules were selected randomly from a set of drug-like compounds first published in or after 2015 (see Supplementary Information for the entire list of the targets and routes).

In the first test, the participants had the choice between a route reported by expert chemists in the literature, and a route generated by our 3N-MCTS algorithm for the same target molecule. Routes to nine different target molecules were offered. Routes towards the same molecule were required to have the same number of steps.

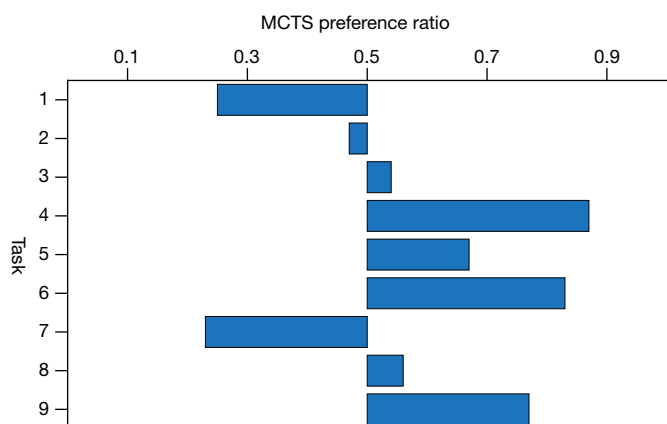
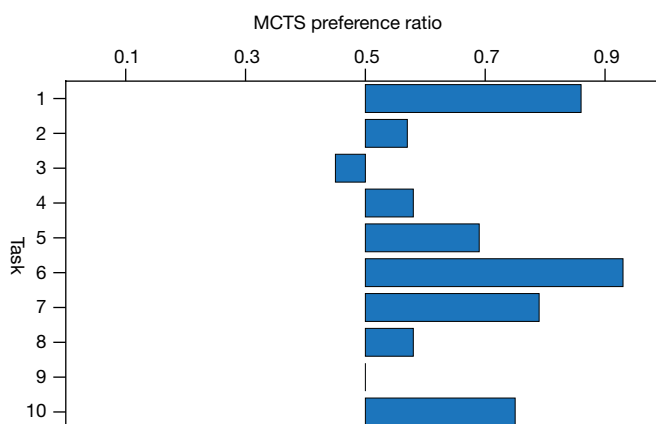
Here, one might expect the participants to clearly identify the routes suggested by the machine as inferior. Surprisingly, this is not the case. We found that the experts did not significantly ($P = 0.26$) prefer the literature route (43.0%) over our program's route (57.0%). Figure 5a shows the preference ratios for the individual routes. Here, the preference is generally balanced, with a slight trend towards MCTS. In some cases, the participants have clear preferences (see Fig. 5c and Extended Data Fig. 3 for examples where MCTS was not preferred).

Table 1 | Experimental results

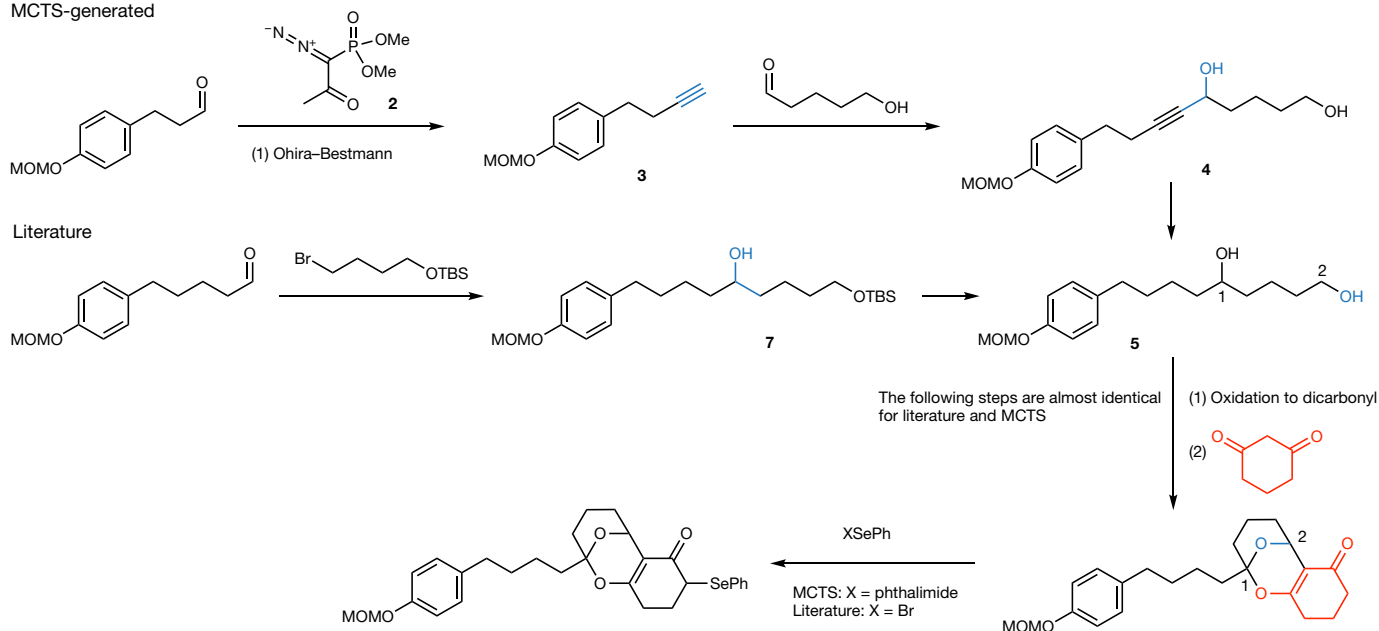
Entry	Search method	Policy*	Percentage solved	Time (seconds per molecule)
1	MCTS	Neural	95.24 ± 0.09	13.0
2	MC	Neural	89.54 ± 0.59	275.7
3	UCT	Neural	87.12 ± 0.29	30.4
4	BFS	Neural	84.24 ± 0.09	39.1
5	BFS	SMILES ^{3/2}	55.53 ± 2.02	422.1

The time budget was 300 s and 100,000 iterations for MCTS or 300 s and 100,000 expansions for BFS, per molecule. Three restarts were carried out.

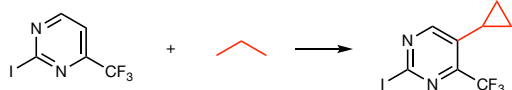
*In the BFS, this is the cost function.

a 3N-MCTS versus literature routes (test a)**b** 3N-MCTS versus heuristic BFS routes (test b)**c** Why did chemists prefer the literature over MCTS in task 1 of test a?

MCTS-generated

**d** Problematic steps in heuristic BFS (without expansion policy and in-scope filter) in test b

Task 6



Task 10

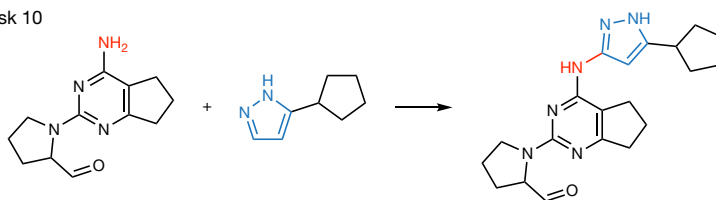


Figure 5 | Double-blind AB testing of MCTS-derived routes against literature and BFS routes. **a**, Chemists did not significantly prefer literature routes over routes found by MCTS (Wilcoxon signed-rank test, $P = 0.26$). **b**, Chemists significantly prefer routes found by 3N-MCTS over routes generated by heuristic BFS without a policy network and an in-scope filter (Wilcoxon signed-rank test, $P = 0.01$). A ratio above 0.5 indicates that more than 50% of participants preferred the MCTS solution. **c**, In this AB example (task 1 of test a), the chemists preferred the literature route proceeding via a Grignard reaction, in contrast to the MCTS route,

which was proposed to proceed via Seyferth–Gilbert homologation with the Ohira–Bestmann reagent. Although the MCTS route is chemically reasonable, it uses less-conventional chemistry in this case. The subsequent key steps to build the annulated cycle are the same for both MCTS and the literature. **d**, Without applying the expansion policy and the in-scope filter to select the best reactions, heuristic BFS produces the typical errors traditionally criticized in retrosynthetic systems. That is, the expert system tries to apply rules that are overgeneral and will not work in this molecular context.

In the second test, the participants had to report their preferences for either routes found by 3N-MCTS or routes generated by a baseline system, which uses heuristic BFS and the same transformation rules as 3N-MCTS. However, it lacks a policy network to

preselect promising transformations and an in-scope filter to exclude unlikely steps. Here, the participants significantly ($P = 0.01$) preferred the routes generated by the MCTS algorithm (68.2%) over the baseline system (31.8%). We attribute the preference towards the

3N-MCTS-generated routes to lower frequencies of unreasonable steps (see Fig. 5b and d).

Discussion

We have shown that MCTS combined with deep neural networks and symbolic rules can be used effectively to perform chemical synthesis planning. In contrast to earlier work, our purely data-driven approach can be initially set up within a few days without the need for tedious and biased expert encoding or curation, and is applicable to discipline-scale datasets. Our approach solves more problems and is faster than established search methods. Furthermore, it also performs better qualitatively. In the past, retrosynthetic systems have been criticized for producing more noise than signal. We observed that traditional heuristic BFS without neural network guiding did lead to many unreasonable steps being proposed in the routes, while the 3N-MCTS approach proposed more reasonable routes. This is supported by double-blind AB experiments, where the participating organic chemists showed clear preference towards 3N-MCTS over the traditional approach. Finally, our double-blind AB tests suggest that, for the first time, organic chemists should consider the quality of retrosynthetic routes generated by a machine to be on par with reported routes for molecules of practical relevance.

Limitations and frontiers

Nevertheless, it would be premature to consider computer-aided synthesis a solved problem, as challenges remain. First, natural product synthesis is currently beyond the capabilities of our method. The sparsity of the training data in this area remains a fundamental challenge for deep learning approaches⁴⁵. However, natural products are also challenging for the best human chemists, as they can behave unpredictably, and often require intense methodology development⁴⁶. Natural product synthesis may be solvable by stronger, but slower-reasoning, algorithms that could be used to invent reactions^{41,47}.

Another important challenge is the reliable prediction of stereochemical outcomes. While our approach is able to treat stereo-information, the most important part, predicting enantiomeric or diastereomeric ratios quantitatively, remains an open challenge. Convincing global approaches for the quantitative prediction of enantiomeric or diastereomeric ratios over a wide range of different reactions without recourse to expensive quantum-mechanical calculations⁴⁸ have not been reported. However, they could be addressed with stereochemistry-aware descriptors. Furthermore, our system currently does not take reaction mechanisms, equilibria between different forms, such as tautomers, or three-dimensional structures into account, which can be crucial in natural product synthesis. Also, we do not at present perform reaction condition prediction⁴⁹.

Outlook

For the past 60 years, experts have been trying to dictate the rules of chemistry to computers via hand-coded heuristics. Instead, we anticipate that equipping machines with strong, general planning algorithms, symbolic representations, and the means to learn autonomously from the rich history of chemistry will be crucial to allowing the machine to become accepted as a valuable assistant in chemical synthesis, which is central to solving humanity's most pressing problems in agriculture, healthcare and material science.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August 2017; accepted 31 January 2018.

1. Clayden, J., Greeves, N., Warren, S. & Wothers, P. *Organic Chemistry* 2nd edn (Oxford Univ. Press, 2008).
2. Brückner, R. *Reaktionsmechanismen: Organische Reaktionen, Stereochemie, Moderne Synthesemethoden* (Springer, 2014).
3. Robinson, R. LXIII. A synthesis of tropinone. *J. Chem. Soc. Trans.* **111**, 762–768 (1917).
4. Corey, E. & Cheng, X. *The Logic of Chemical Synthesis* (Wiley, 1989).
5. Kurti, L. & Czako, B. *Strategic Applications of Named Reactions in Organic Synthesis* (Elsevier, 2005).
6. Evans, J. in *The Oxford Handbook of Thinking and Reasoning* (eds Holyoak, K. J. & Morrison, R. G.) 115–133 (Oxford Univ. Press, 2012).
7. Collins, K. D. & Glorius, F. A robustness screen for the rapid assessment of chemical reactions. *Nat. Chem.* **5**, 597–601 (2013).
8. Ley, S. V., Fitzpatrick, D. E., Ingham, R. & Myers, R. M. Organic synthesis: march of the machines. *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015).
9. Schneider, P. & Schneider, G. De novo design at the edge of chaos: miniperspective. *J. Med. Chem.* **59**, 4077–4086 (2016).
10. Segler, M. H., Kogej, T., Tyrchan, C. & Waller, M. P. Generating focussed molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **4**, 120–131 (2018).
11. Vléduts, G. Concerning one system of classification and codification of organic reactions. *Inform. Storage Retrieval* **1**, 117–146 (1963).
12. Todd, M. H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **34**, 247–266 (2005).
13. Szymkuć, S. et al. Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).
14. Cook, A. et al. Computer-aided synthesis design: 40 years on. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 79–107 (2012).
15. Ihlenfeldt, W.-D. & Gasteiger, J. Computer-assisted planning of organic syntheses: the second generation of programs. *Angew. Chem. Int. Edn Engl.* **34**, 2613–2633 (1996).
16. Fick, R. *Konzepte zur Syntheseplanung: Strukturelle Ähnlichkeit und Strategische Bindungen*. PhD thesis, Friedrich-Alexander-Universität (1996).
17. Ugi, I. et al. Models, concepts, theories, and formal languages in chemistry and their use as a basis for computer assistance in chemistry. *J. Chem. Inf. Comput. Sci.* **34**, 3–16 (1994).
18. Kayala, M. A., Azencott, C.-A., Chen, J. H. & Baldi, P. Learning to predict chemical reactions. *J. Chem. Inf. Model.* **51**, 2209–2222 (2011).
19. Minsky, M. A *Framework for Representing Knowledge*. Technical Report (Massachusetts Institute of Technology, 1974).
20. Bøgevig, A. et al. Route design in the 21st century: the ICSYNTH software tool as an idea generator for synthesis prediction. *Org. Process Res. Dev.* **19**, 357–368 (2015).
21. Law, J. et al. Route designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation. *J. Chem. Inf. Model.* **49**, 593–602 (2009).
22. Christ, C. D., Zentgraf, M. & Kriegl, J. M. Mining electronic laboratory notebooks: analysis, retrosynthesis, and reaction based enumeration. *J. Chem. Inf. Model.* **52**, 1745–1756 (2012).
23. Segler, M. H. & Waller, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry* **23**, 5966–5971 (2017).
24. Boda, K., Seidel, T. & Gasteiger, J. Structure and reaction based evaluation of synthetic accessibility. *J. Comput. Aided Mol. Des.* **21**, 311–325 (2007).
25. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **1**, 8 (2009).
26. Coulom, R. Efficient selectivity and backup operators in Monte-Carlo tree search. In *Int. Conf. on Computers and Games* 72–83 (Springer, 2006).
27. Kocsis, L. & Szepesvári, C. Bandit based Monte-Carlo planning. In *17th Eur. Conf. on Machine Learning* 282–293 (Springer, 2006).
28. Browne, C. B. et al. A survey of Monte Carlo tree search methods. *IEEE Trans. Comput. Intell. AI Games* **4**, 1–43 (2012).
29. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (MIT Press, in the press).
30. Coulom, R. Computing “elo ratings” of move patterns in the game of go. *ICGA J.* **30**, 198–208 (2007).
31. Stern, D., Herbrich, R. & Graepel, T. Bayesian pattern ranking for move prediction in the game of Go. In *Int. Conf. on Machine Learning* 873–880 (Omni Press, 2006).
32. Maddison, C. J., Huang, A., Sutskever, I. & Silver, D. Move evaluation in Go using deep convolutional neural networks. In *3rd Int. Conf. on Learning Representations* (2015); preprint at <https://arxiv.org/abs/1412.6564>.
33. Clark, C. & Storkey, A. Training deep convolutional neural networks to play Go. In *32nd Int. Conf. on Machine Learning* 1766–1774 (PMLR, 2015); <http://proceedings.mlr.press/v37/clark15.html>.
34. Winands, M. Neural networks for video game AI. In *Artificial and Computational Intelligence in Games: Integration (Dagstuhl Seminar 15051)* Vol. 5 (eds Lucas, S. M. et al.) 224 (2015).
35. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
36. Reaxys <http://www.reaxys.com> (Elsevier Life Sciences, 2017).
37. Srivastava, R. K., Greff, K. & Schmidhuber, J. Training very deep networks. In *Advances in Neural Information Processing Systems* 2377–2385 (MIT Press, 2015); preprint at <https://arxiv.org/abs/1507.06228>.
38. Clevert, D.-A., Unterthiner, T. & Hochreiter, S. Fast and accurate deep network learning by exponential linear units (ELUs). In *4th Int. Conf. on Learning Representations* (2016); preprint at <https://arxiv.org/abs/1511.07289>.
39. Sheridan, R. P. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J. Chem. Inf. Model.* **53**, 783–790 (2013).

40. Marcou, G. *et al.* Expert system for predicting reaction conditions: the Michael reaction case. *J. Chem. Inf. Model.* **55**, 239–250 (2015).
41. Segler, M. H. & Waller, M. P. Modelling chemical reasoning to predict and invent reactions. *Chemistry* **23**, 6118–6128 (2017).
42. Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H. & Jensen, K. F. Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.* **3**, 434–443 (2017).
43. Murphy, K. P. *Machine Learning: a Probabilistic Perspective* (MIT Press, 2012).
44. Nirogi, R. V., Badange, R., Reballi, V. & Khagga, M. Design, synthesis and biological evaluation of novel benzopyran sulfonamide derivatives as 5-HT₆ receptor ligands. *Asian J. Chem.* **27**, 2117–2124 (2015).
45. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, 1–101 (2016).
46. Sierra, M. A. & de la Torre, M. C. Dead ends and detours en route to total syntheses of the 1990s. *Angew. Chem. Int. Ed.* **39**, 1538–1559 (2000).
47. Rocktäschel, T. & Riedel, S. End-to-end differentiable proving. In *Advances of Neural Information Processing Systems* (eds Guyon, I. *et al.*) 3788–3800 (Curran Associates, 2017); <https://papers.nips.cc/paper/6969-end-to-end-differentiable-proving>.
48. Peng, Q., Duarte, F. & Paton, R. S. Computing organic stereoselectivity—from concepts to quantitative calculations and predictions. *Chem. Soc. Rev.* **45**, 6093–6107 (2016).
49. Lin, A. I. *et al.* Automatized assessment of protective group reactivity: a step toward big reaction data analysis. *J. Chem. Inf. Model.* **56**, 2140–2148 (2016).
50. Gini, A., Segler, M., Kellner, D. & Garcia Mancheno, O. Dehydrogenative tempo-mediated formation of unstable nitrones: easy access to n-carbamoyl isoxazolines. *Chemistry* **21**, 12053–12060 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.H.S.S. and M.P.W. thank the Deutsche Forschungsgemeinschaft (SFB858) for funding. M.H.S.S. and M.P.W. also thank D. Evans (RELX Intellectual Properties) and J. Swienty-Busch (Elsevier Information Systems) for the reaction dataset. We thank all AB-test participants in Shanghai and Münster, and J. Guo for assistance in AB testing. M.H.S.S. thanks M. Wiesenfeldt, the Studer group, D. Barton, S. McAnanama-Brereton, R. Vidyadharan and T. Kogej for discussions. M.P. thanks M. Winands and J. Togelius for insights.

Author Contributions M.H.S.S. conceived the project, M.P.W. and M.P. contributed ideas. M.H.S.S., M.P. and M.P.W. designed the experiments. M.H.S.S. implemented the program. M.H.S.S. and M.P.W. conducted the experiments. M.P.W. supervised the project. All authors co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.H.S.S. (marwin.segler@www.de) or M.P.W. (waller@shu.edu.cn).

Reviewer Information *Nature* thanks D. Duvenaud, W. H. Green and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Chemistry. Molecules are stored in the search tree as canonical SMILES strings. If stereochemistry is desired, molecules are stored as canonical, isomeric SMILES. For processing, molecules are translated from SMILES into molecular graphs, which are vertex-labelled and edge-labelled graphs $m = (A, B)$, with atoms $a \in A$ as vertices and bonds $b \in B$ as edges. Retrosynthetic transformation rules are productions on graphs⁵¹. In chemical terminology, transformations are also referred to as 'named reactions'. The Chemistry Development Kit (CDK)⁵² and RDKit⁵³ cheminformatics libraries were used for the implementation.

Retrosynthesis as a Markov decision process. Markov decision processes (MDPs) model sequential decision processes of an agent in an environment²⁹. An MDP is a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R})$, with states (positions) $s \in \mathcal{S}$, actions (transformations) $a \in \mathcal{A}$, a transition model $\mathcal{T}(s, a, s')$ determining the probability $\Pr(s'|a, s)$ of reaching state s' when taking action a in state s , and a reward function $\mathcal{R}(s, a, s')$, which returns the reward when transitioning to s' via action a in state s . A policy $\pi(a|s)$ is a probability distribution over all actions given state s .

Unlike in games, such as chess or Go, where it is trivial to write down the ground truth rules (the model) of the game, querying the 'chemical environment' to find out whether an action actually leads to the desired successor state is expensive. Either a wet-laboratory experiment has to be conducted or a quantum-chemical calculation on a high level of theory has to be run, which usually takes longer than running the laboratory experiment. Learning from millions of episodes of self-play can therefore not be employed.

To avoid these expensive interactions, we therefore need to learn or construct a model of the environment to perform planning. As elaborated in the introduction, this model will be inaccurate^{29,54}. Even the best human chemists' predictions can and do fail, which implies that humans also perform synthesis planning with inaccurate mental models of the chemical environment. Here, we use automatic rule extraction to determine the action set (the transformations) and the in-scope filter network to learn a transition model (which is applied in a binary way). The expansion policy network serves as a prior policy.

In this Article, a state (position) $s \in \mathcal{S}$ is a set of molecules $s = \{m_i, m_j, \dots\}$. The initial state $s_0 = \{m_0\}$ contains only the target molecule m_0 . Actions are then transformations (rules or productions on graphs⁵¹) applied to one of the molecules m in a state s . When applying a legal action a_i to a state $s_i = \{m_a, m_b, m_c\}$, it will produce a new state, for example, $s_j = \{m_d, m_e, m_b, m_c\}$.

Given a set of building-block molecules \mathcal{B} (specified before the start of the search), state s_k is solved if all molecules m_i in s_k are building blocks $m_i \in \mathcal{B}$. A state is terminal either if it is solved or if no legal actions are available. Ideally, it should be possible to provide the set of building blocks \mathcal{B} dynamically before the search is started. Otherwise, the system could not adapt, for example, when a building block runs out of stock. Also, a researcher may choose different sets of building blocks for each search, for example, by first trying to find the solution to a problem with molecules that are in stock in the laboratory, and afterwards considering additional molecules from chemical suppliers. This makes it challenging to define or learn value functions, because a changed set of building blocks changes the terminality of states and the reward function, which entails a change of the value function²⁹. A further challenge is that the initial state (the target molecule) changes. In most games, the reward function is always the same (the rules never change, and a terminal state is always terminal).

The size of the state space can only be roughly estimated. The number of drug-like molecules, which contain a restricted set of elements and functional groups, might already exceed 10^{60} molecules⁵⁵. However, this number excludes synthetic intermediates and organometallic and organo-main group chemistry, which add orders of magnitude to the state space size. The action space is formed by the transformations available to the system, and the legal actions are those actions that can be applied to the molecules in a state via subgraph isomorphism. Unlike for other game artificial intelligence problems, in retrosynthesis we can limit the depth of the tree to a relatively small number (here, 25) and abort the simulation as failed if a viable route cannot be found within this limit. Our trees are thus wider and less deep than for other applications.

MCTS. MCTS is a reinforcement learning approach that combines tree search with learning from simulated episodes of experience, which are obtained from interacting with a model of the environment^{28,29}. MCTS has been successfully applied in problems of sequential decision making in many domains, such as games or automated theorem proving^{28,56}. MCTS has several desirable features, which makes it particularly well suited for retrosynthesis. It allows the calculation of value functions focused on a particular initial state on the fly⁵⁴, and therefore does not depend strongly on heuristics. Each edge (s, a) in the search tree stores the action value $Q(s, a)$, the visit count $N(s, a)$ and a prior probability $P(s, a)$ received from the expansion policy network.

Selection. In the first MCTS phase, starting at the root node, the tree policy (equation (1)) is used to select actions. The simulation descends the tree step by

step. At each step t , the next action a_t is selected from all available actions $\mathcal{A}(s_t)$ in s_t by equation (1), where $N(s_{t-1}, a_{t-1})$ is the visit count of the state-action pair that led to the current state, and c the exploration constant.

$$a_t = \operatorname{argmax}_{a \in \mathcal{A}(s_t)} \left(\frac{Q(s_t, a)}{N(s_t, a)} + cP(s_t, a) \frac{\sqrt{N(s_{t-1}, a_{t-1})}}{1 + N(s_t, a)} \right) \quad (1)$$

The inclusion of the prior probability $P(s, a)$ in the second term of equation (1) allows the system to explore the most promising lines of analysis first⁵⁷. With repeated visits this term decays, allowing for the exploration of other options. Additionally, this allows one to take into account the confidence in the evaluation obtained via the rollout, which is expected to be noisy.

The tree policy is applied until a leaf node or a terminal node is found. If a leaf node is visited for the second time, it is expanded. Then, all non-building-block molecules $m_i \in s_t$ are processed by means of the expansion procedure. The resulting state-action pairs are added to the tree as children of the leaf node, and the most probable action according to the policy network is selected for rollout.

Expansion. During expansion, the state is processed once via the expansion procedure, and the reduced top 50 successor states are directly added to the tree. This trick can be applied because retrosynthesis is a single-player game, and we do not have to fear overlooking 'killer' moves (trap moves in which a small mistake will be exploited directly by the opponent) as much as in two-player games. Using only the reduced top 50 entails that the NP-complete subgraph isomorphism problem, which determines whether the corresponding rule can be applied to a molecule and yields the next molecule(s), needs only to be solved for at most 50 rules, instead of for all rules in the transformation rule set.

Evaluation by rollout. Before starting the rollout, the state is first checked for being terminal. A state can be terminal if it is solved. States within the tree that are already solved are called 'proved'⁵⁸. States can also be terminal if no legal actions are available in that state. Terminal states are directly evaluated with the reward function. If the state is non-terminal, a rollout is started. During rollout, actions are sampled recursively for each molecule in the state from the top 10 actions of the rollout policy until it has been deconstructed into building blocks or a maximal recursive function call depth of d_r is exceeded. For the sake of simplicity, rollouts that completely solve the molecule are currently not stored.

The reward function $r(s)$ returns $z > 1$ if the state is proved to encourage exploitation, a reward $\in [0, 1]$ depending on the ratio of molecules solved during rollout, and -1 if the state is terminal and unproved, or unsolved during rollout. Learned value functions are a possible, future alternative to rollouts. Investigations to learn value functions are currently ongoing in our laboratory.

Update. In the update phase, the action values $Q(s, a)$ and visit counts $N(s, a)$ of the edges traversed in the branch from s_t to the root node are updated. The edges gather the mean action value as in equation (2), where the indicator function $I_i(s, a)$ is 1 if the edge was played during the i th simulation and z_i is the reward received during rollout.

$$Q(s, a) = \frac{1}{N(s, a)} \sum_{i=1}^n I_i(s, a) z_i W(b_i) \quad (2)$$

Here, it is also possible to inject custom objective functions $W(b_i) \rightarrow \mathbb{R}$ that might assign higher rewards, for example, to shorter, convergent, atom-economic or confident branches b_i . In this work, we adjust the reward by using

$$\xi(b_i) = \text{length}(b_i) - \sum_{j=1}^J kP(s_j, a_j) \quad (3)$$

$$W^{\text{Lmax}}(b_i) = \max \left(0, \frac{L_{\text{max}} - \xi(b_i)}{L_{\text{max}}} \right) \quad (4)$$

where $J = \text{length}(\cdot)$ denotes the length of a branch, $P(s_j, a_j)$ is the probability of the j th action in the branch obtained from the expansion policy, $k = 0.99$ is a damping factor, and the maximal branch length $L_{\text{max}} = 25$. Here, inclusion of the prior policy $P(s_j, a_j)$ allows us to bias the reward also towards more confident branches. An interesting scoring function to investigate further is similarity to reported reactions^{59,60}.

After either the time or the iteration step budget has been exhausted, the synthesis plan is selected, starting at the root node, by greedily choosing the action with the highest action value until a terminal solved state is reached, or a maximum depth has been exceeded, in which case the problem is unsolved. A maximal rollout depth of 5, an exploration constant of 3 and a reward, when proved, of 10 were employed as the MCTS parameters in the quantitative and qualitative experiments.

Automatic transformation rule extraction. Formalizing chemical knowledge by hand has been attempted. Even though it sounds simple to write down the rules of chemistry, it takes years to formalize only humble knowledge bases, it is error-prone and biased towards the knowledge of the encoding experts, and in many cases chemical systems are too complex or just not well enough understood to formally write down their limitations and scope. Like rule-based common-sense reasoning¹⁹, this approach is considered to have exhausted its potential^{15,17,18}. Additionally, given the exponential growth of chemical knowledge (it doubles roughly every 15 years), manual encoding is a hopeless endeavour.

Following our previously reported procedure²³, and building on previous work^{20–22}, transformation rules were therefore extracted automatically. The rules are stored using the RDKit reaction SMARTS format⁵³. A very general rule definition was employed for the expansion rule set, where only the atoms of the reaction centre (including implicit hydrogen atoms and neighbouring-atom count) were extracted. The rules in the rollout set contain the reaction centre atoms (with implicit hydrogen atoms and neighbouring-atom count) and additionally the directly neighbouring atoms of the atoms in the reaction centre with their implicit hydrogen atom count. Rules were extracted only from single-step reactions with one, two or three reactants and a single product. As this work is a proof-of-concept study with the intent to radically avoid expert encoding and curation, we chose not to exclude reactions based on low yield or extreme reaction conditions, as these are quite subjective criteria. For example, a yield of just a few per cent can be sufficient if the aim is to obtain only a few milligrams of a compound for biological testing, while 90% yield is clearly unsatisfying if quantitative alternatives are available. In the future, more sophisticated approaches based on reaction classification could be employed to extract rules, for example, by grouping together similar leaving groups into a single rule^{21,61,62}. Also, further investigation of directly translating from products to reactants using neural networks is called for⁶³.

The general advantages and limitations of automatic rule extraction have been discussed in detail elsewhere^{13–15,20,21,23}. Its main disadvantages (defining the scope of reactions and competing reactivity, incorporating mechanistically needed/activating functional groups, and deciding which rules to apply first) can be addressed by learning supervised policies to predict which rules to apply²³.

The use of symbolic rules has the great advantage that it is deeply rooted in chemists' language. This makes it easy for the model to communicate its results to the human user. Furthermore, because the transformations were extracted from the literature, we can link back directly to literature precedents, which is of crucial importance for chemists.

We note that even when taking all pre-2015 rules without count restriction into account, only 82% of the reactions published in and after 2015 are covered. The missing 18% are novel reaction types. This highlights the success of chemists inventing novel methodologies, but also implies that eventually a retrosynthesis system should also be able to discover novel reactions on its own²³.

Policy networks. The neural policy networks were trained by minimizing the negative log-likelihood of selecting the transformation a that was used in the literature to make molecule m . This is essentially supervised multi-class classification. To evaluate the accuracy, reactions that are not covered in the rule set are excluded. Training was carried out using stochastic gradient descent (ADAM optimizer⁶⁴) within 1–2 days on a single NVIDIA K80 graphics processing unit. The Keras neural network framework was employed, using Theano as the backend^{65,66}.

Expansion policy network. Molecules are represented by real vectors in the form of counted extended-connectivity fingerprints (ECFP4)⁶⁷, which are first modulo-folded to 1,000,000 dimensions, and then $\ln(x + 1)$ -preprocessed. After that, a variance threshold is applied to remove rare features, leaving 32,681 dimensions. For the machine-learning model, we used a 1+5-layer highway network with exponential linear unit (ELU) nonlinearities^{37,38}. A dropout ratio of 0.3 was applied after the first affine transformation to 512 dimensions, and a dropout ratio of 0.1 was applied after each of the five highway layers. The last layer of the neural networks is a softmax, which outputs a probability distribution over all actions (transformations) $p(a|m)$, which forms the policy (see Extended Data Fig. 4).

Rollout policy network. For rollout, molecules are represented by counted ECFP4 fingerprints⁶⁷, modulo-folded to 8,192 dimensions, and then $\ln(x + 1)$ -preprocessed. As the rollout policy, a neural network with a single hidden dense layer with a dimensionality of 512, and ELU nonlinearity and dropout of 0.4 was used.

In-scope filter. The function of the in-scope classifier is to rapidly filter out the nonsensical reactions that plague rule-based systems, such as incorrect regioisomers in electrophilic aromatic substitutions. For this purpose, we chose a binary classifier, which is fast to evaluate. The investigation of more sophisticated, but slower, reaction prediction approaches is left to future work^{18,23,41,42,68–70}. For the same reason, only the product and the reaction fingerprint serve as inputs to the classifier, although the exclusion of conditions makes the classifier underdetermined^{41,42}. Reactions can selectively lead to different products under different

conditions. However, the inclusion of reaction conditions as another input feature would require additional search in condition space at each step, which is not feasible given the time constraints we imposed here. In tasks where search time is not a constraint, such as in process development, reaction condition prediction could be performed at each step of the search, or be performed in a second sweep. One way to predict reaction conditions might be via reaction similarity search, which is related to how chemists use reaction databases.

Our classifier is a neural network with two branches (see Extended Data Fig. 4). The first branch embeds the reaction r_i , represented as a counted ECFP4 reaction fingerprint^{40,59,71–73} ρ_i modulo-folded to 2,048 dimensions, via a single dense ELU layer. The second branch embeds the product, represented as a counted ECFP4 fingerprint φ_i , modulo-folded to 16,384 dimensions and $\ln(x + 1)$ -preprocessed, through a 5-layer ELU-highway network. The cosine proximity of these embeddings is then fed into a sigmoid unit to predict the probability that the reaction gives rise to the expected product.

We used two strategies to obtain negative data, as follows. First, 30 million incorrect reactions were obtained by the application of reaction rules to the reactants of reported reactions, using the same rule set as for the expansion policy^{41,42,74}. Here we make the assumption that the reactants can only react in the reported way. Any product generated by rule application not matching the reported product is considered to be a failed product. With this approach, for example, wrong regioisomers can be generated. We note that these 'negative' reactions generated in this way capture the cases where a naive, contextless rule-based system would fail. Furthermore, 70 million negative training data points were generated by perturbing tuples (ρ_i, φ_i) to (ρ_j, φ_j) , where $i \neq j$, by random sampling. Random sampling gives a small performance boost of 0.0025 score points in the area under the receiver operation characteristic curve (AUROC) and 0.0072 points in the area under the precision-recall curve. Training data were generated only from reactions published before 2015, while test data were generated from data published in or after 2015. The classifier was trained by minimizing negative log-likelihood using the ADAM optimizer. Extended Data Fig. 1 shows the receiver operation characteristic curve of the classifier. A value of 0.9 was selected as the decision threshold for our classifier. Here, the classifier has a false negative rate of 14% and a false positive rate of 1.5%.

Given that the in-scope filter learns to embed molecules in a vector space close to the reactions that were used to make them, it would be interesting to investigate if this can be directly used for nearest-neighbour search of molecules in reaction-rule space, which can also be described as label embedding.

Rediscovering electronic properties. To study what the in-scope filter has learned, we conducted two experiments. First, we studied Diels–Alder reactions of cyclopentadiene with various dienophiles. The reactions were submitted to the in-scope filter, and the raw logit scores (the output of the neural network before applying the final sigmoid function) were calculated. As a comparison, the structures of the dienophiles were optimized and the energies of the LUMO were calculated via density functional theory (BP86-D3BJ/def2-SVP) to capture the qualitative trend, using the ORCA3 software⁷⁵. Extended Data Fig. 5a shows the correlation of the LUMO energy with the logit score, which has an $r^2 = 0.74$. Additionally, para-brominations via electrophilic aromatic substitution were studied (Extended Data Fig. 5b). Here, the logit scores were correlated to Hammond electronic parameters, and an $r^2 = 0.78$ was found. This indicates that the in-scope filter correlates with basic electronic properties, following the expected behaviour of the respective reaction mechanisms. This is remarkable, as its input features (ECFP4-based molecular and reaction fingerprints) do not contain electronic information.

Performance evaluation studies. Baselines. In BFS, each branch is added to a priority queue, which is sorted by cost. In heuristic BFS, this cost function is the SMILES^{3/2} heuristic, which is used as reported¹³. In neural BFS, the cost is calculated as $f(b) = \sum_{i=0}^{s=b} (1 - P(a_i|s_i))$, where $P(a_i|s_i)$ is the probability of that transformation calculated by the expansion policy. Evaluation studies were performed using the central processing unit of a 24-core commodity cluster node using a single search thread. To provide a more meaningful comparison no graphical processing unit was used for the evaluation studies.

Building block and test molecule selection. The building blocks have to be selected before the search is started, and could be molecules on stock in the laboratory, known in the literature, or commercially available chemicals. Here we use a set of 423,731 molecules, containing 84,253 building blocks from three major chemical suppliers (SigmaAldrich, AlfaAesar and Acros), obtained from the ZINC database (<http://zinc15.docking.org/>) and 339,478 molecules from the Reaxys database, which have been used as reactants at least five times before 2015. To obtain a set of target molecules for the quantitative evaluation that contains different scaffolds, first all molecules reported after 2014 were clustered using the ECFP6-based Butina algorithm⁷⁶. Then, 497 target molecules were randomly selected from amongst the 82,673 different cluster cores.

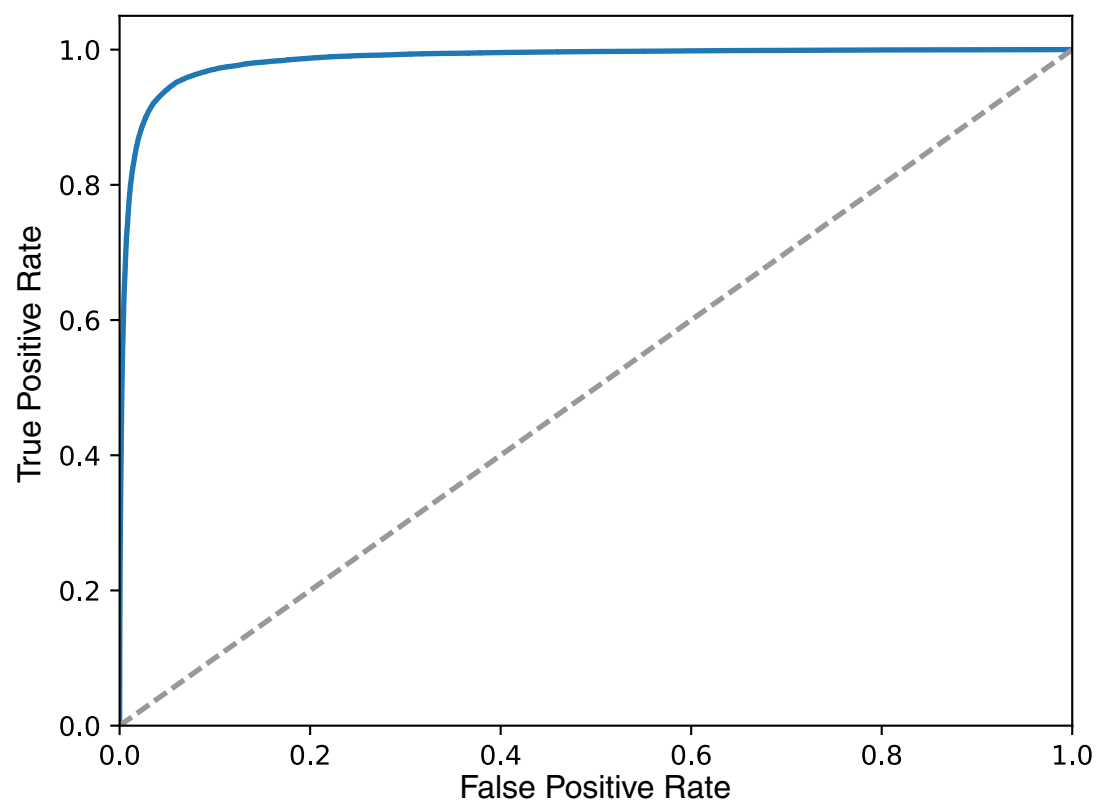
AB testing. The participants in the AB tests were 45 postgraduate students who had specialized in organic chemistry at the Institute of Organic Chemistry at Westfälische Wilhelms-Universität Münster and Shanghai Institute of Organic Chemistry. The study was conducted in a double-blind setup. During the test, neither the participants nor the conductors were aware of the origin of the route. Statistical significance was tested via the Wilcoxon signed-rank test.

3N-MCTS versus literature. In the comparison of 3N-MCTS with the literature, the expectation would be that experts prefer the literature option. In 128 AB tests, the experts preferred the literature route in 43.0% and MCTS in 57.0% (Wilcoxon signed-rank test on paired data, $P = 0.26$). The null hypothesis that both datasets stem from the same source cannot be rejected.

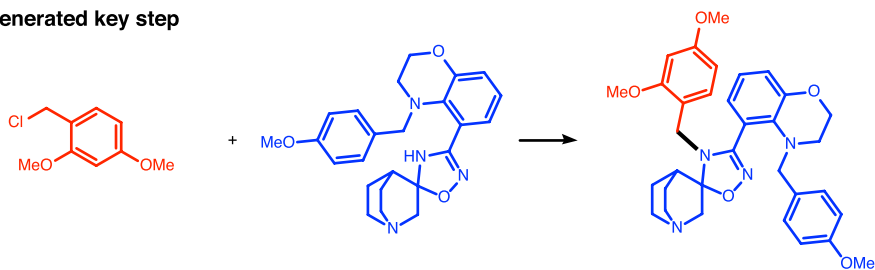
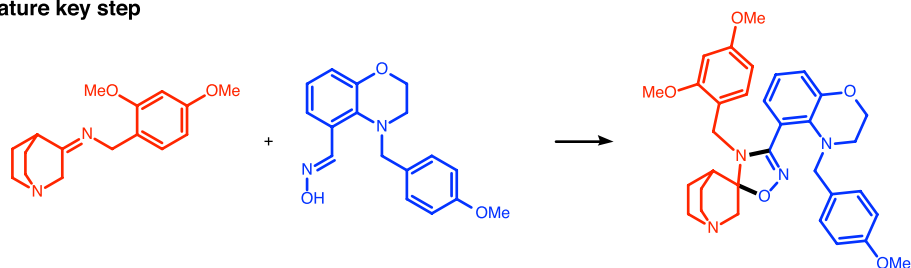
3N-MCTS versus heuristic BFS. Here, 68.2% of the participants preferred 3N-MCTS generated solutions, whereas only 31.8% preferred heuristic-BFS-generated solutions in 129 submitted tests. The experts strongly favour MCTS, the null hypothesis of indistinguishable sources (50% preference for each) can clearly be rejected (Wilcoxon signed-rank test on paired data, $P = 0.01277$).

Data availability. The reaction dataset used in this study is provided by Elsevier Information Systems GmbH under licence.

51. Andersen, J. L., Flamm, C., Merkle, D. & Stadler, P. F. Generic strategies for chemical space exploration. *Int. J. Comput. Biol. Drug Des.* **7**, 225–258 (2014).
52. Steinbeck, C. *et al.* Recent developments of the chemistry development kit (CDK)—an open-source Java library for chemo- and bioinformatics. *Curr. Pharm. Des.* **12**, 2111–2120 (2006).
53. Landrum, G. *RDKit: Open-Source Cheminformatics* <http://www.rdkit.org>.
54. Silver, D. *Reinforcement Learning and Simulation-Based Search*. PhD thesis, Univ. Alberta (2009).
55. Raymond, J.-L., Ruddigkeit, L., Blum, L. & van Deursen, R. The enumeration of chemical space. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 717–733 (2012).
56. Färber, M., Kaliszyk, C. & Urban, J. Monte Carlo connection prover. Preprint at <https://arxiv.org/abs/1611.05990> (2016).
57. Rosin, C. D. Multi-armed bandits with episode context. *Ann. Math. Artif. Intell.* **61**, 203–230 (2011).
58. Winands, M. H., Björnsson, Y. & Saito, J.-T. Monte-Carlo tree search solver. In *Int. Conf. on Computers and Games* 25–36 (Springer, 2008).
59. Schneider, N., Lowe, D. M., Sayle, R. A. & Landrum, G. A. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *J. Chem. Inf. Mod.* **55**, (2015).
60. Coley, C. W., Rogers, L., Green, W. H. & Jensen, K. F. Computer-assisted retrosynthesis based on molecular similarity. *ACS Cent. Sci.* **3**, 1237–1245 (2017).
61. Gelernter, H., Rose, J. R. & Chen, C. Building and refining a knowledge base for synthetic organic chemistry via the methodology of inductive and deductive machine learning. *J. Chem. Inf. Comput. Sci.* **30**, 492–504 (1990).
62. Rose, J. R. & Gasteiger, J. Horace: an automatic system for the hierarchical classification of chemical reactions. *J. Chem. Inf. Comput. Sci.* **34**, 74–90 (1994).
63. Liu, B. *et al.* Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent. Sci.* **3**, 1103–1113 (2017).
64. Kingma, D. P. & Ba, J. ADAM: a method for stochastic optimization. In *3rd Int. Conf. for Learning Representations*; preprint at <https://arxiv.org/abs/1412.6980> (2015).
65. Chollet, F. *et al.* Keras <https://github.com/fchollet/keras> (2015).
66. The Theano Development Team Theano: a Python framework for fast computation of mathematical expressions. Preprint at <https://arxiv.org/abs/1605.02688> (2016).
67. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
68. Wei, J. N., Duvenaud, D. & Aspuru-Guzik, A. Neural networks for the prediction of organic chemistry reactions. *ACS Cent. Sci.* **2**, 725–732 (2016).
69. Socorro, I. M. & Goodman, J. M. The ROBIA program for predicting organic reactivity. *J. Chem. Inf. Model.* **46**, 606–614 (2006).
70. Satoh, H. & Funatsu, K. Sophia, a knowledge base-guided reaction prediction system—utilization of a knowledge base derived from a reaction database. *J. Chem. Inf. Comput. Sci.* **35**, 34–44 (1995).
71. Patel, H., Bodkin, M. J., Chen, B. & Gillet, V. J. Knowledge-based approach to de novo design using reaction vectors. *J. Chem. Inf. Model.* **49**, 1163–1184 (2009).
72. Zhang, Q.-Y. & Aires-de Sousa, J. Structure-based classification of chemical reactions without assignment of reaction centers. *J. Chem. Inf. Model.* **45**, 1775–1783 (2005).
73. Polishchuk, P. *et al.* Structure–reactivity modeling using mixture-based representation of chemical reactions. *J. Comput. Aided Mol. Des.* **31**, 829–839 (2017).
74. Carrera, G. V., Gupta, S. & Aires-de Sousa, J. Machine learning of chemical reactivity from databases of organic reactions. *J. Comput. Aided Mol. Des.* **23**, 419–429 (2009).
75. Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2**, 73–78 (2012).
76. Butina, D. Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **39**, 747–750 (1999).
77. Parsy, C. C. *et al.* Discovery and structural diversity of the hepatitis C virus NS3/4a serine protease inhibitor series leading to clinical candidate IDX320. *Bioorg. Med. Chem. Lett.* **25**, 5427–5436 (2015).



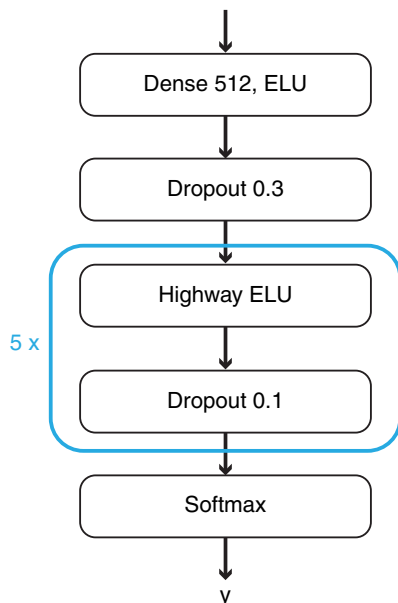
Extended Data Figure 1 | Receiver operation characteristic curve for the in-scope filter. The area under the curve is 0.99.

a) Why did chemists prefer the literature over MCTS in Test a) Task 7?**MCTS-Generated key step****Literature key step**

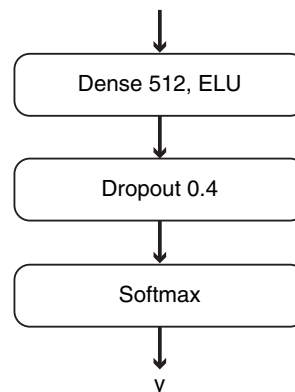
Extended Data Figure 3 | Example of reaction used in the AB testing, where the MCTS-derived route was less favoured. In this task, the participants preferred the literature solution, as its key step was presumably perceived to be more convergent.

Expansion Policy

Product Fingerprint
ECFP4, folded to 1,000,000 dim; $\log(x+1)$
VarianceThreshold >> 32681 dim

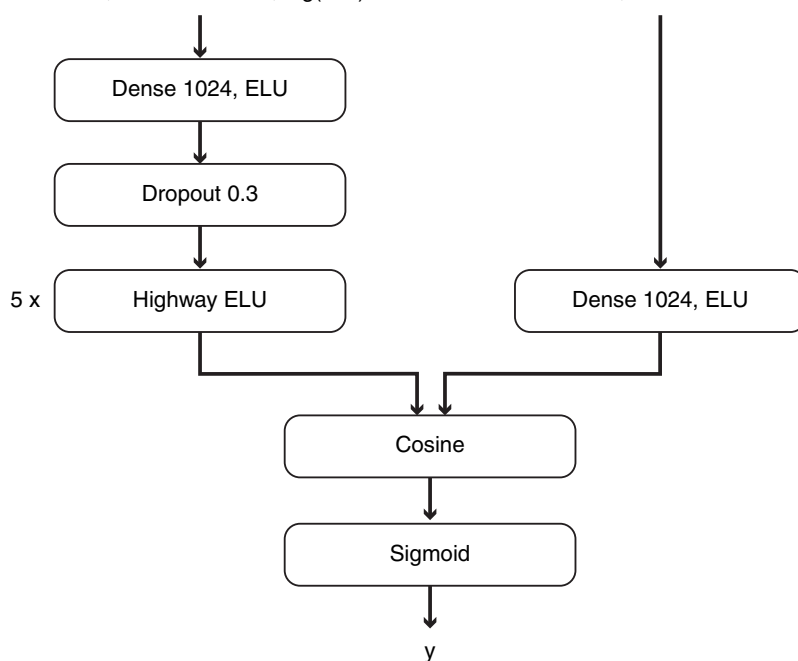
**Rollout Policy**

Product Fingerprint
ECFP4, folded to 8192 dim; $\log(x+1)$

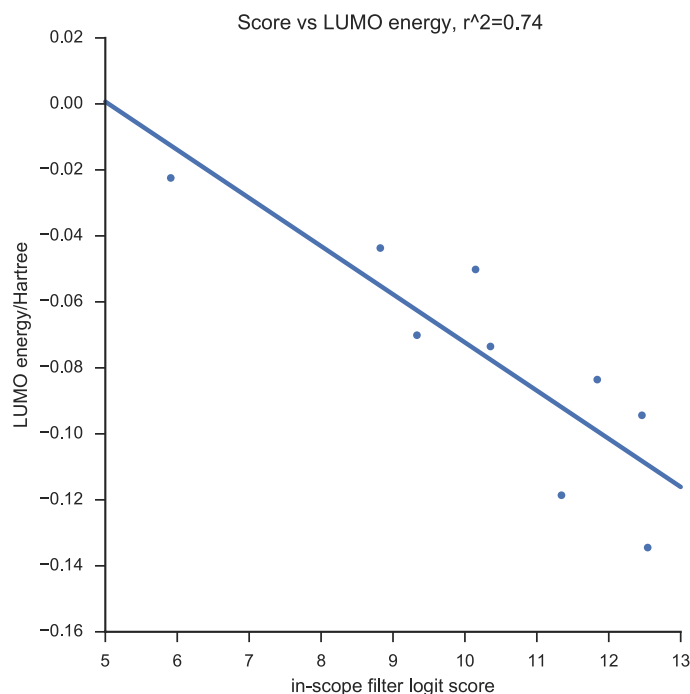
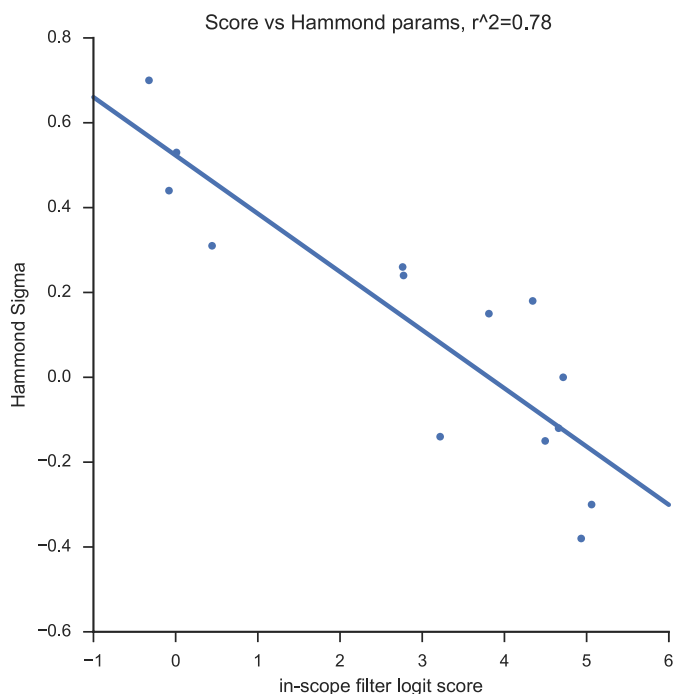
**In Scope Filter**

Product Fingerprint
ECFP4, folded to 16384, $\log(x+1)$

Reaction Fingerprint
ECFP4, folded to 2048



Extended Data Figure 4 | Architectures of the employed neural networks. ('dim', dimensions.)

a) Diels-Alder reactions with Cyclopentadiene**b) para-Bromination of benzenes**

Extended Data Figure 5 | Rediscovering physicochemical properties with the in-scope filter. The output logit score of the neural network correlates surprisingly well with calculated quantum-mechanical properties (LUMO energies, in Hartree) in Diels–Alder reactions

($r^2 = 0.74$) (a) and with empirically measured Hammond parameters in electrophilic brominations ($r^2 = 0.78$) (b), even though the input features (ECFP4 fingerprints) do not contain electronic information.

Extended Data Table 1 | Metrics for the supervised neural network policies

Policy	# rules	Coverage	Matching rules/mol ^b	Accuracy ^a	top10Acc ^a	top50Acc ^a
Expansion	301,671	0.79	46,175	0.310	0.633	0.725
Rollout	17,134	0.52	321	0.501	0.891	0.964

Top10Acc/top50Acc is the ratio of correct/incorrect predictions if we allow the system to make 10 or 50 predictions.

^aAccuracy is calculated on the molecules covered by the respective rulebase.

^bMatching rules/mol corresponds to the branching factor.

De novo mutations in regulatory elements in neurodevelopmental disorders

Patrick J. Short¹, Jeremy F. McRae¹, Giuseppe Gallone¹, Alejandro Sifrim¹, Hyejung Won², Daniel H. Geschwind^{2,3,4}, Caroline F. Wright^{1,5}, Helen V. Firth^{1,6}, David R. FitzPatrick^{1,7}, Jeffrey C. Barrett¹ & Matthew E. Hurles¹

We previously estimated that 42% of patients with severe developmental disorders carry pathogenic *de novo* mutations in coding sequences. The role of *de novo* mutations in regulatory elements affecting genes associated with developmental disorders, or other genes, has been essentially unexplored. We identified *de novo* mutations in three classes of putative regulatory elements in almost 8,000 patients with developmental disorders. Here we show that *de novo* mutations in highly evolutionarily conserved fetal brain-active elements are significantly and specifically enriched in neurodevelopmental disorders. We identified a significant twofold enrichment of recurrently mutated elements. We estimate that, genome-wide, 1–3% of patients without a diagnostic coding variant carry pathogenic *de novo* mutations in fetal brain-active regulatory elements and that only 0.15% of all possible mutations within highly conserved fetal brain-active elements cause neurodevelopmental disorders with a dominant mechanism. Our findings represent a robust estimate of the contribution of *de novo* mutations in regulatory elements to this genetically heterogeneous set of disorders, and emphasize the importance of combining functional and evolutionary evidence to identify regulatory causes of genetic disorders.

The importance of non-coding variation in complex disease has been well established—most disease-associated common SNPs lie in intergenic or intronic regions, albeit with low effect sizes^{1,2}. Rare sequence and structural variants in relatively few regulatory elements have been causally linked to Mendelian disorders^{3–5}. These pathogenic regulatory variants can act by loss of function^{6–9} or gain of function^{10,11} and most act dominantly, with a few exceptions¹². These regulatory elements can lie far from the gene they regulate. For example, sequence variants in an evolutionarily conserved regulatory element located 1 Mb from its target gene, *SHH*, can cause polydactyly¹⁰. As a consequence, it can be challenging to identify the gene whose regulation is being perturbed by an associated regulatory variant^{13–15}. Moreover, the contribution of highly penetrant mutations in regulatory elements to genetically heterogeneous rare diseases, such as neurodevelopmental disorders, has not been firmly established.

We recruited 7,930 individuals with a severe, undiagnosed developmental disorder, and their parents to the Deciphering Developmental Disorders (DDD) study from clinical genetics centres in the UK and Ireland. Systematic clinical phenotyping¹⁶ identified 79% with cognitive impairment or abnormality of the brain, which we refer to as neurodevelopmental disorders. Congenital heart defects (CHD) were the most prevalent non-neurodevelopmental phenotype, present in 10% of the cohort. Exome sequencing of the first 4,293 families in a previous analysis revealed that about 25% of probands carry damaging *de novo* mutations (DNMs) in genes associated with developmental disorders, accounting for the majority of diagnostic variants^{17,18}. An additional 17% of probands carry pathogenic DNMs in genes not yet robustly associated with developmental disorders¹⁸. Thus the majority of the probands do not carry a diagnostic variant in a protein-coding gene, and are termed ‘exome-negative’. To explore the role of DNMs in non-coding elements, we performed targeted sequencing on three classes of putative regulatory elements: 4,307 highly evolutionarily

conserved non-coding elements (CNEs)¹⁹, 595 experimentally validated enhancers²⁰, and 1,237 putative heart enhancers²¹, together covering 4.2 Mb of sequence with comparable depth of coverage to protein-coding regions (Extended Data Fig. 1, Supplementary Table 1). Furthermore, we define a set of ‘control’ intronic elements covering 6.03 Mb (see Methods).

Selective constraint acting on non-coding elements

We first assessed how much purifying selection had skewed allele frequencies in non-coding elements. We used the mutability-adjusted proportion of singletons (MAPS) metric²² in 7,080 unrelated, unaffected DDD parents to test six different element classes: introns, heart enhancers, validated enhancers, CNEs, protein-coding genes, and genes known to be associated with developmental disorders. The validated enhancers from the VISTA enhancer browser vary across the spectrum of evolutionary conservation, while the heart enhancers are poorly conserved, consistent with previous reports²³, and the CNEs show high levels of evolutionary conservation (Fig. 1a). The introns and heart enhancers show little evidence of purifying selection, while the experimentally validated enhancers and CNEs are constrained to a similar degree to protein-coding genes, but less than genes known to be associated with developmental disorders (Fig. 1b), consistent with evolutionary conservation maintained by purifying selection. Statistical power to detect functionally relevant variants in protein-coding genes is strengthened considerably by stratification of variants by their likely impact on the encoded protein and variant deleteriousness metrics such as CADD²⁴. We computed the MAPS within bins of CADD scores encompassing 1,520,250 variants in unaffected DDD parents to assess whether CADD was predictive of selective constraint. In protein-coding genes, the strong correlation between CADD score and strength of purifying selection enabled us to differentiate between variants that are neutral, weakly constrained, and highly constrained.

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. ³Center for Autism Research and Treatment, Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. ⁴Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. ⁵Institute of Biomedical and Clinical Science, University of Exeter Medical School, RILD Level 4, Royal Devon & Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK. ⁶East Anglian Medical Genetics Service, Box 134, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. ⁷MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK.

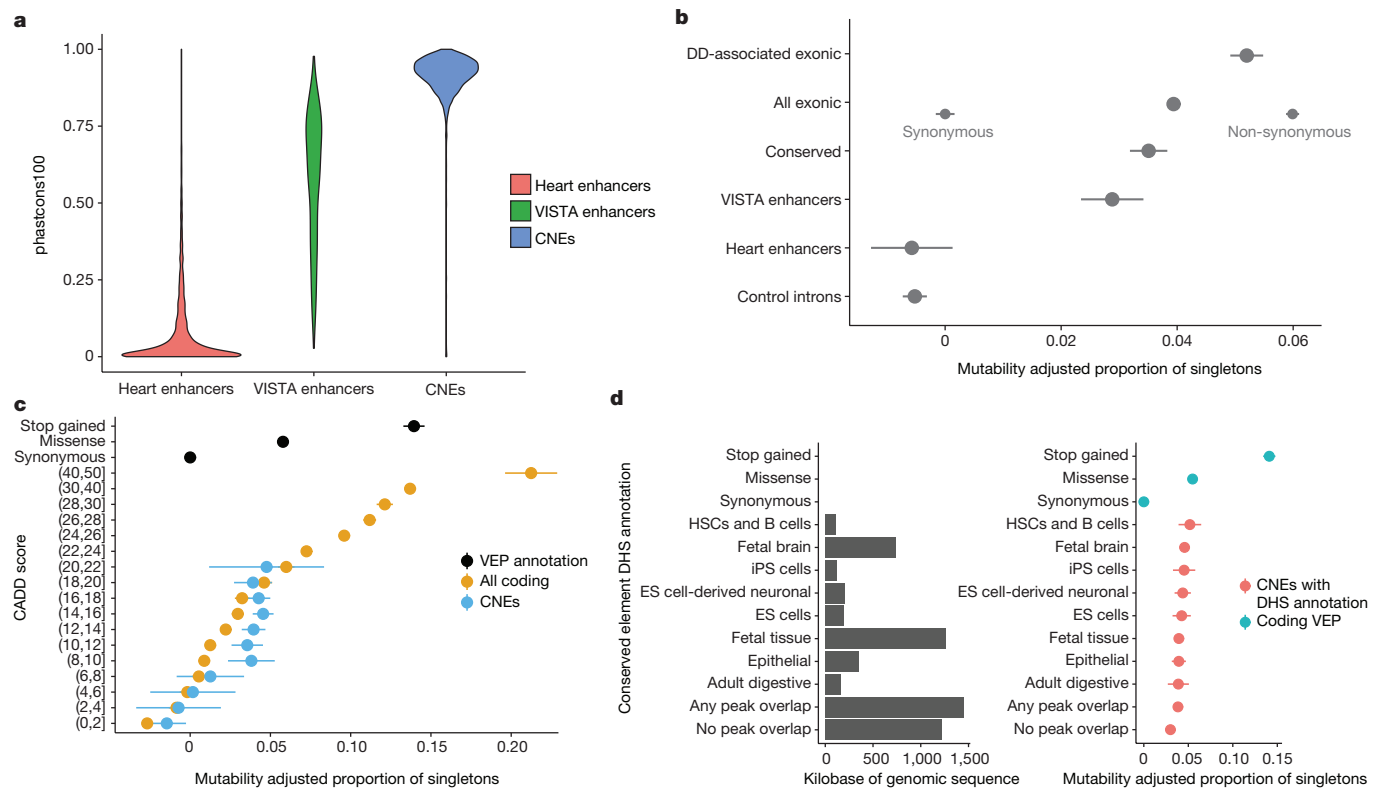


Figure 1 | Selective constraint in targeted non-coding elements.

a, Evolutionary conservation score (phastcons100) for CNEs ($n = 4,307$), experimentally validated enhancers (VISTA; $n = 595$), and putative heart enhancers ($n = 1,237$). **b**, Strength of selection (MAPS metric, mean and 95% CI represented by dot and bars) in targeted non-coding elements compared to protein-coding regions, where 'Exonic' refers to all variation within protein coding-exons. Stratification based on synonymous/non-synonymous consequence displayed on the same row to illustrate power of even a simple discriminator. Introns and putative heart enhancers show little evidence of purifying selection while CNEs show

In CNEs, CADD differentiates neutral variation from variation under weak constraint, but failed to identify highly deleterious variants with selective constraint on a par with protein-truncating variants (Fig. 1c, Extended Data Fig. 2d). Other deleteriousness metrics were assessed, but none were more informative than CADD (Extended Data Fig. 2a–c).

We used DNase I hypersensitivity sites (DHS) in 39 tissues and chromHMM genome segmentation predictions in 111 tissues²⁵ to predict tissue activity for the targeted non-coding elements. Of the 4,307 CNEs we sequenced, 4,046 (93.9%) were active in at least one of the 111 surveyed tissues whereas 261 (6.1%) were inactive or repressed in all tissues (Extended Data Fig. 2e, f). Variants within a DHS peak in at least one tissue were under stronger purifying selection than variants that did not overlap a DHS peak ($P = 0.019$), but we did not identify significant differences in selective constraint between tissues (Fig. 1d).

Enrichment of mutations in non-coding elements

We identified candidate *de novo* single nucleotide mutations in 7,930 trios (see Methods). We adapted a previously described model for germline mutation²⁶ to include methylation status at CpG sites (see Methods, Extended Data Fig. 3a) and show that it better accounts for observed levels of rare variation than the unadapted model (Extended Data Fig. 3b). We tested four genomic features previously associated with mutagenicity²⁷ for enrichment in non-coding elements with DNMs and found no evidence that these genomic features were enriched in non-coding elements with DNMs (H3K27me3, χ^2 -test $P = 0.4809$; H3K9me3, χ^2 -test $P = 0.1966$; replication timing²⁸, Extended Data Fig. 3f; recombination rate²⁹, Extended Data Fig. 3e).

selection on par with all genes, but less than genes known to be associated with developmental disorders. **c**, Using CADD to stratify coding and non-coding variants observed in unaffected parents differentiates neutral variation from weakly and strongly constrained sites in coding regions, but fails to identify non-coding variation with selection pressure on par with protein-truncating variants (stop gained). **d**, Sites overlapping a DHS in at least one tissue are under stronger purifying selection than sites not overlapping a DHS. ES cells, embryonic stem cells; HSCs, haematopoietic stem cells; iPS cells, induced pluripotent stem cells.

We identified 1,691 'exome-positive' individuals with a likely pathogenic protein-altering DNM or inherited variant in a gene known to be associated with developmental disorders, with the remaining 6,239 being 'exome-negative'. Using the mutation model, we compared the numbers of observed and expected DNMs in the targeted non-coding elements in these individuals. No significant DNM enrichment was observed in exome-positive probands in the targeted non-coding elements, demonstrating that the mutation model is reasonably well-calibrated and that a large proportion of exome-positive cases are likely to represent Mendelian syndromes caused by high-penetrance protein-coding mutations (Extended Data Fig. 4a). We note that the number of exome-positive individuals affords only limited power to reject modest mutation enrichment in the non-coding elements. On the basis of these results, we chose to focus on the 6,239 exome-negative individuals for subsequent analyses.

We found that the CNEs were nominally significantly enriched for DNMs (422 observed, 388 expected, $P = 0.04$), whereas experimentally validated enhancers (153 observed, 156 expected, $P = 0.605$), heart enhancers (86 observed, 86 expected, $P = 0.514$), and intronic controls (901 observed, 919 expected, $P = 0.728$) were not enriched (Fig. 2).

Given the preponderance of individuals with neurodevelopmental disorders in our cohort but the broad range of tissue activity of the targeted CNEs, we focused on CNEs that are active in the fetal brain. DNMs were strongly and significantly enriched within 2,077 fetal brain DHS peaks in CNEs (177 observed, 138 expected, $P = 8.1 \times 10^{-4}$) but no enrichment in sites in CNEs falling outside fetal brain DHSs (245 observed, 249 expected, $P = 0.608$) (Fig. 2). We also used chromHMM³⁰ predictions of fetal brain activity and

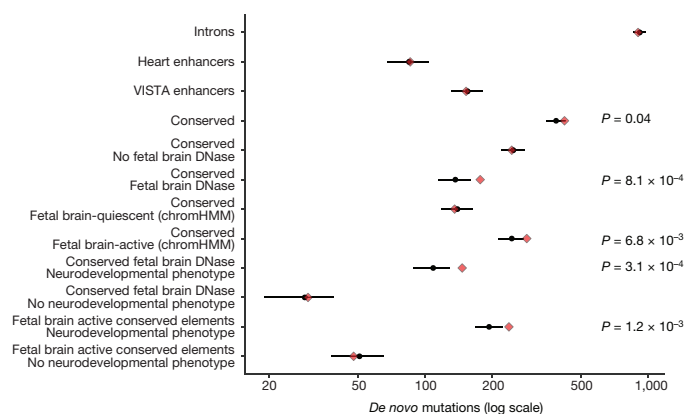


Figure 2 | Enrichment of DNMs across element classes and functional annotations in exome-negative probands. $n = 6,239$. Red diamonds indicate observed counts, while black circles and bars indicate expected count and 95% CI, respectively. Targeted CNEs showed a modest enrichment for DNMs (422 observed, 388 expected, $P = 0.04$) while heart enhancers, experimentally validated enhancers, and control introns matched the null model. Observed enrichment is specific to CNEs predicted to be active in the fetal brain and to patients with neurodevelopmental disorders (238 observed, 194 expected, $P = 1.2 \times 10^{-3}$). Confidence intervals and P values derived from a Poisson distribution.

again identified significant enrichment of DNMs in the 2,613 fetal brain-active CNEs (Fig. 2). Moreover, the DNMs observed in fetal brain-active CNEs in exome-negative probands were at more highly conserved sites (Wilcoxon rank sum test on PhyloP 100-way score³¹) compared to DNMs observed in exome-positive probands (Extended Data Fig. 4b). To test for as yet unknown factors causing differential mutability, we compared the levels of rare variation in fetal brain-active and -inactive CNEs in 7,509 deep whole genomes from the gnomAD consortium and found no evidence for a higher germline mutation rate in fetal brain-active elements (Extended Data Fig. 3c, d). The excess of DNMs observed in fetal brain-active CNEs is concentrated exclusively within the 79% of exome-negative probands with neurodevelopmental phenotypes (fetal brain DHS peaks: 147 observed, 109 expected, $P = 3.1 \times 10^{-4}$; fetal brain-active by chromHMM: 238 observed, 194 expected, $P = 1.2 \times 10^{-3}$), with no significant enrichment observed in those without neurodevelopmental phenotypes (fetal brain DHS: $P = 0.413$; fetal brain-active by chromHMM: $P = 0.681$) (Fig. 2). The highly significant and specific enrichment of DNMs in fetal brain-active CNEs in exome-negative probands with neurodevelopmental disorders is robust to Bonferroni correction for thirteen explicitly and implicitly tested hypotheses (see Methods, Extended Data Fig. 5a). Analysis of the FANTOM5³² and EnhancerAtlas³³ datasets suggests that 50–70% of the fetal brain-active CNEs act as enhancers (see Methods).

We re-evaluated the experimentally validated enhancers with functional evidence for activity in fetal brain ($N = 383$, 64%) and observed a nominally significant enrichment for DNMs only within the top quartile of evolutionary conservation (18 observed, 9 expected, $P = 0.01$) (Extended Data Fig. 5b). This result suggests that even for experimentally validated fetal brain enhancers, DNM enrichment is concentrated within elements with strong evolutionary conservation.

We assessed four methods of gene target prediction: Genomicus¹⁴ (based on evolutionary synteny), correlation between DNase accessibility and gene expression³⁴, Hi-C in fetal brain¹⁵ and choosing the closest gene. Genome annotations are rapidly evolving and the sensitivity and specificity of gene target prediction methods is not yet known. However, independent expression quantitative trait loci, enhancer RNA and Hi-C data all suggest that the closest gene is often not the target of non-coding regulatory variation^{32–34}.

Across the four methods tested, the proportion of fetal brain-active CNEs for which a target gene was predicted was 28% (fetal brain Hi-C),

48% (DHS-RNA correlation), 91% (evolutionary synteny), and 100% (closest gene). The pairwise concordance between any two methods (given that both methods make a prediction) was between 17% and 35% (Extended Data Fig. 6a). Intersecting multiple independent methods may provide higher confidence predictions, but comes at a cost of sensitivity and therefore power. We did not identify any enrichment for DNMs in elements predicted to target genes known to be associated with developmental disorders, likely dosage-sensitive genes (pLI metric²²), or genes that are differentially expressed in the brain (see Methods, Extended Data Fig. 6b for Hi-C results). Elements with DNMs were enriched for interactions with genes that are specifically upregulated in early prenatal brain development³⁵ (Extended Data Fig. 6c, Methods).

We assessed the impact of DNMs on a set of 45 transcription factor binding motifs that are enriched in fetal brain-active CNEs (see Methods), and observed a nominally significant enrichment for DNMs predicted to increase binding affinity; this did not survive multiple hypothesis correction (Extended Data Fig. 7a–d). Given the number of DNMs we have identified, and the relative immaturity of *in silico* predictions of the impact of non-coding variation, it is not currently possible to determine precise mechanisms by which these DNMs contribute to developmental disorders.

To explore the penetrance associated with the observed DNM enrichment in the targeted non-coding elements, we investigated potential overtransmission of inherited rare variants in these elements to affected children and found no evidence for overtransmission (Extended Data Fig. 7e). Furthermore, we did not detect any enrichment for rare variants *in cis* that would suggest that the DNM is acting as a ‘second hit’ to an already perturbed haplotype. The fold-enrichment of DNMs is consistent with DNMs in fetal brain-active CNEs comprising a mixture of 70–80% non-pathogenic DNMs and 20–30% pathogenic DNMs.

Recurrently mutated regulatory elements

We found a significant excess of recurrently mutated elements (two or more DNMs in unrelated individuals) in the fetal brain-active CNEs and evolutionarily conserved enhancers compared to the expectation under the null mutation model (31 observed, 15 expected, $P = 9.3 \times 10^{-5}$) (Fig. 3a). However, no individual element exceeded a conservative genome-wide significance threshold of $P < 1.91 \times 10^{-5}$ (Bonferroni correction for independent tests on 2,613 fetal brain-active elements) (Fig. 3b).

Increased power to detect locus-specific enrichments of DNMs could be gained from aggregating DNMs across elements that regulate the same target gene. However, as described above, gene target prediction lacks coverage and accuracy. CNEs have been shown to cluster together within the genome and are enriched around developmentally important genes³⁶. Therefore we applied hierarchical clustering on the 2,613 fetal brain-active CNEs to identify 356 clusters (see Methods). We found an excess of recurrently mutated clusters, defined as two or more elements with at least one DNM in each element (11 observed, 6 expected, $P = 0.016$). We did not find any element clusters with a significant excess of DNMs at a genome-wide significance threshold (Supplementary Table 2).

We used chromHMM³⁰ to assign the recurrently mutated CNEs to a predicted chromatin state. We observed the greatest excess of DNMs in CNEs predicted to be enhancers ($n = 9$) or strongly or weakly transcribed ($n = 8$) (Extended Data Fig. 8). Five of the eight transcribed recurrently mutated elements fall in close proximity to exons, but are not in protein-coding transcripts and show evidence of involvement in alternative splicing (*BCLAF1*, *SRRT*, *SLC10A7*, and *MKNK1*) or as a 3′ UTR (*CELFI*). The full set of recurrently mutated elements is described in Supplementary Table 3 and the location of DNMs relative to population variation and additional annotations is shown in Extended Data Fig. 9.

Estimating genome-wide non-coding mutation burden

The absence of individual non-coding elements with genome-wide significant enrichment of DNMs allowed us to place an upper bound on

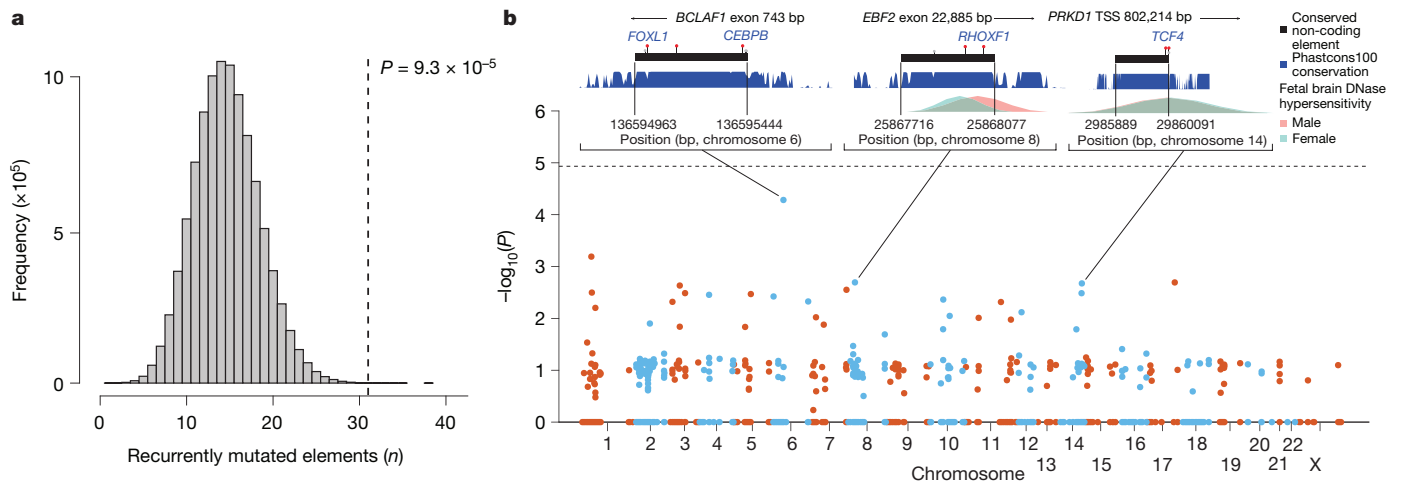


Figure 3 | Recurrently mutated elements. **a**, Approximately twofold enrichment of recurrently mutated non-coding elements. Grey histogram shows distribution of expected number of recurrently mutated fetal brain-active non-coding elements under the null model and vertical line indicates observed number. **b**, Enrichment test of individual non-coding elements. No element was significant at a genome-wide threshold of $P < 1.9 \times 10^{-5}$ (Bonferroni correction for testing 2,613 fetal brain-

active elements). Inset plots for three elements show the nearest exon or transcription start site, location of DNMs (red markers) with any predicted transcription factor binding site disruptions (gain of binding in blue, loss of binding in red), location of rare variants in unaffected parents (grey markers), evolutionary conservation (blue, higher indicates more conserved), and fetal brain DNase I hypersensitivity (male in pink, female in blue). TSS, transcription start site.

the proportion of sites and elements in which DNMs are pathogenic. Approximately 8% of DNMs in protein-coding regions result in a protein-truncating mutation^{26,37}. CNEs are smaller than protein-coding exons (median 600 bp) and also lack annotation to identify putative pathogenic mutations. Down-sampling gene length to 600 bp and masking protein consequence annotation resulted in an 80% drop in empirical power for the 94 genes passing the genome-wide significance threshold in a previous study¹⁸ (Extended Data Fig. 10a). As we did not discover any genome-wide significant CNEs, the proportion of DNMs in CNEs that are pathogenic and highly penetrant must be substantially lower than 8%. We modelled the likelihood of observing 286 DNMs, 25 recurrently mutated CNEs, and zero CNEs at genome-wide significance across different values for the number of fetal brain-active CNEs (out of 2,613) and the proportion of mutations in those elements that are pathogenic with a dominant mechanism for neurodevelopmental disorders (see Methods). The maximum likelihood model is one in which 3.5% of mutations within approximately 100 elements are pathogenic with a dominant mechanism. However, there is considerable uncertainty around this point estimate (Extended Data Fig. 10b), with the credible interval including scenarios in which tens of elements have around 5–7% of mutations being pathogenic or thousands of elements have below 1% of mutations being pathogenic.

Our survey of the non-coding genome is biased towards highly evolutionarily conserved elements, but also includes elements with lower levels of evolutionary conservation. To extrapolate the excess of DNMs we observed in the targeted non-coding elements to a genome-wide estimate, we modelled the enrichment of DNMs as a function of evolutionary conservation (see Methods). Factoring in the distribution of evolutionary conservation of fetal brain DHS peaks genome-wide, we predicted a genome-wide excess of 88 DNMs (95% confidence interval (CI): 48–140), corresponding to 1.0–2.8% of exome-negative cases carrying pathogenic mutations in regulatory elements (Fig. 4b) in contrast to 13.4% and 28.4% carrying protein-truncating variants and missense variants, respectively, estimated previously¹⁸ (Fig. 4c).

Discussion

We have demonstrated that *de novo* mutations in regulatory elements contribute to severe neurodevelopmental disorders. These elements act primarily either as enhancers or to regulate alternative splicing, but establishing the precise mechanism for each element has proved challenging. This significant excess of DNMs is observed only in highly

evolutionarily conserved elements that are active in the fetal brain. These elements also exhibit substantial selective constraint within human populations. We observed a 1.3-fold excess of DNMs within DHS peaks in these regulatory elements, suggesting that a minority of such DNMs are pathogenic. Moreover, our modelling suggests that there are few, if any, regulatory elements in which more than 4% of mutations cause neurodevelopmental disorders with a dominant mechanism. Our data are consistent with only 0.15% of mutations within fetal brain-active CNEs being highly penetrant for neurodevelopmental disorders (Fig. 4a); this is likely to be considerably lower than the proportion of dominant pathogenic mutations in protein-coding regions. As a consequence, this class of pathogenic non-coding DNMs is likely to account for only a small proportion (less than 5%) of ‘exome-negative’ individuals, and the robust identification of disease-associated regulatory elements will present a greater challenge than of protein-coding genes.

Our study design focuses on highly conserved elements and fetal brain-active elements, and is relatively uninformative with respect to pathogenic ‘gain-of-function’ DNMs within elements that show no wild-type activity in fetal brain and are not highly evolutionarily conserved. While our findings have focused on the highly conserved elements, we do not consider our observations to be definitively negative about the role of less highly conserved fetal brain enhancers in neurodevelopmental disorders, or the role of heart enhancers in CHD (owing to the low proportion of subjects with CHD). The field of regulatory element annotation has progressed tremendously over the six years since this study design was initially conceived. Therefore, a comprehensive analysis of the contribution of variation within all classes of non-coding elements to neurodevelopmental disorders is likely to require whole genome sequencing (WGS) of many tens of thousands, if not hundreds of thousands, of parent–proband trios (Extended Data Fig. 10c).

One challenge of interpreting WGS data is the vast universe of hypotheses that could be tested, and thus how to account appropriately for multiple hypothesis testing. A recent study reported a nominally significant enrichment ($P = 0.03$) of *de novo* single-nucleotide variants (SNVs) and private copy number variants in fetal brain DHS or at sites with PhyloP conservation scores above 4, within 50 kb of known autism-associated genes in WGS from 53 individuals with autism³⁸. Caution should be exercised in interpreting findings based on small sample sizes relative to those required for well-powered analyses (as discussed above) and analyses requiring multiple, arbitrary levels of

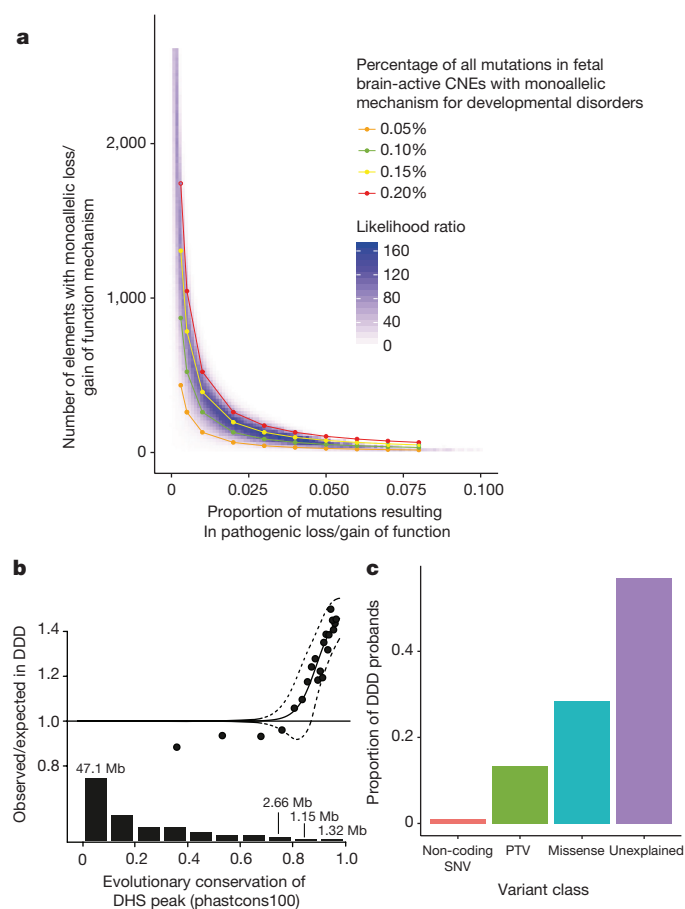


Figure 4 | Modelling the proportion of DNMs in non-coding elements that are likely to be highly penetrant for dominant neurodevelopmental disorders. **a**, Our observation of zero non-coding elements at genome-wide significance in 6,239 exome-negative probands indicates that very few sites within these elements (<5%) are likely to contribute to developmental disorders through a highly penetrant dominant mechanism. **b**, Logistic regression used to model the genome-wide contribution of dominant-acting DNMs in fetal brain DNase hypersensitive sites in non-coding elements as a function of level of evolutionary conservation using a sliding window approach including 1,000 elements in each bin (see Methods). Dashed lines indicate the upper and lower 95% CI. The bar plot shows fetal brain-active DHS peaks genome-wide (in megabase of total sequence) at a given level of evolutionary conservation. **c**, The proportion of probands carrying a pathogenic de novo SNV in a fetal brain-active regulatory element (1–2.8%) is far lower than the proportion carrying a pathogenic protein-truncating DNM (~13.4%) or missense DNM (~28.4%).

variant stratification (for example, gene set, genomic proximity threshold, and conservation score). WGS-based analyses need to account for all explicit and implicit hypothesis testing.

Our analyses were limited to SNVs as current mutation models for indels and structural variation are too inaccurate to allow robust assessment of mutational excess. In addition, our analyses highlight an urgent need for improved tools to stratify benign and damaging variants within non-coding elements and to annotate gene targets for regulatory elements. These improved mutational models and functionally relevant annotations will greatly increase power to detect highly-penetrant disease-associated non-coding variation, for example, increasing power more than tenfold from 8% to 83% in 40,000 trios (Extended Data Fig. 10c). Functional characterization of increasing numbers of robustly associated, highly-penetrant, regulatory variants in cellular and animal models will be critical in moving from a descriptive to a more predictive understanding of non-coding variation in the human genome, as well as elucidating its underlying pathophysiological mechanisms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February 2017; accepted 24 January 2018.

Published online 21 March 2018.

- Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
- Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
- Mathelier, A., Shi, W. & Wasserman, W. W. Identification of altered cis-regulatory elements in human disease. *Trends Genet.* **31**, 67–76 (2015).
- Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Human Mol. Genet.* **25**, 157–165 (2016).
- Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
- Jeong, Y. *et al.* Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* **40**, 1348–1353 (2008).
- Benko, S. *et al.* Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *J. Med. Genet.* **48**, 825–830 (2011).
- Bhatia, S. *et al.* Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* **93**, 1126–1134 (2013).
- Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).
- Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
- Hill, R. E. & Lettice, L. A. Alterations to the remote control of Shh gene expression cause congenital abnormalities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120357 (2013).
- Sellick, G. S. *et al.* Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat. Genet.* **36**, 1301–1305 (2004).
- Noonan, J. P. & McCallion, A. S. Genomics of long-range regulatory elements. *Annu. Rev. Genomics Hum. Genet.* **11**, 1–23 (2010).
- Naville, M. *et al.* Long-range evolutionary constraints reveal cis-regulatory interactions on the human X chromosome. *Nat. Commun.* **6**, 6904 (2015).
- Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* **48**, 488–496 (2016).
- Köhler, S. *et al.* The Human Phenotype Ontology in 2017. *Nucleic Acids Res.* **45**, D865–D876 (2017).
- Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314 (2015).
- Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
- Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
- Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
- May, D. *et al.* Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* **44**, 89–93 (2011).
- Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Blow, M. J. *et al.* ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810 (2010).
- Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
- Carlson, J. *et al.* Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. Preprint at <https://www.biorxiv.org/content/early/2017/02/14/108290> (2017).
- Koren, A. *et al.* Genetic variation in human DNA replication timing. *Cell* **159**, 1015–1026 (2014).
- Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
- Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
- Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
- Gao, T. *et al.* EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics* **32**, 3543–3551 (2016).
- Shooshtari, P., Huang, H. & Cotsapas, C. Integrative genetic and epigenetic analysis uncovers regulatory mechanisms of autoimmune disease. *Am. J. Hum. Genet.* **101**, 75–86 (2016).

35. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
36. Sandelin, A. *et al.* Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **5**, 99 (2004).
37. Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80**, 727–739 (2007).
38. Turner, T. N. *et al.* Genome sequencing of autism-affected families reveals disruption of putative non-coding regulatory DNA. *Am. J. Hum. Genet.* **98**, 58–74 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the families for their participation and patience; the DDD study clinicians, research nurses and clinical scientists in the recruiting centres for their hard work and perseverance on behalf of families; the Exome Aggregation Consortium and Genome Aggregation Database (<http://gnomad.broadinstitute.org/>) for making their data and code available; S. Gerety, G. Elgar, S. Aerts, and D. Svetlichnyy for discussions; H. Roest Crolius and L. Moyon for help with gene target prediction; J. Mudge and A. Frankish for help in annotating CNEs; and the Sanger HGI and DNA pipelines teams for their support in generating and processing the data. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant HICF-1009-003), a parallel funding partnership between the Wellcome Trust and the UK Department of Health, and the Wellcome Trust Sanger Institute (grant

WT098051). The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the UK Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South Research Ethics Committee and GEN/284/12, granted by the Republic of Ireland Research Ethics Committee). D.R.F. is funded through an MRC Human Genetics Unit program grant to the University of Edinburgh. D.H.G. is funded through 1U01 MH105666 and 1R01 MH110927 (psychENCODE consortium). A.S. is supported by the FWO (Postdoctoral Fellow number 12W7318N).

Author Contributions Study design: H.V.F., C.F.W., D.R.F., J.C.B. and M.E.H. Method development and data analysis: P.J.S., J.F.M., G.G., A.S., H.W., D.H.G., and M.E.H. Writing: P.J.S. and M.E.H. Experimental and analytical supervision: H.V.F., C.F.W., D.R.F., J.C.B. and M.E.H. Project Supervision: M.E.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.E.H. (meh@sanger.ac.uk).

Reviewer Information *Nature* thanks M. Daly and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Defining targeted non-coding elements. The placental mammal 28-way phastCons score¹⁹ was used to select the top 4,432 CNEs with no overlap with RefSeq genes (downloaded from UCSC on 4 August 2010). Using the VISTA enhancer browser²⁰, all 622 putative enhancers with evidence of *in vivo* activity in developing mouse embryos were downloaded on 3 August 2010. At the time the capture was designed, it had been observed that heart enhancers are depleted among ultra-conserved elements²³. As heart defects are the largest group of non-CNS abnormalities in the DDD cohort we sought to supplement the ultra-conserved elements with an early annotation of heart enhancers. These putative heart enhancers were provided by A. Visel and based on chromatin immunoprecipitation with sequencing (ChIP-seq) of p300 in human fetal heart described previously²¹ in GRCh36 coordinates, mapped over to GRCh37. Collectively, these elements cover approximately 4.6 Mb of total sequence. First, elements were filtered to exclude any targeted sequences with less than 10× coverage across the DDD data set. Second, any elements previously annotated to be non-coding, but classified as protein-coding in Gencode v19³⁹, were removed. Finally, any elements less than 50 bp in length were excluded. After filtering, 4,307 conserved elements, 595 enhancer elements and 1,237 putative heart enhancers remained.

Defining intronic control sequences. The exome baits designed to capture the coding regions frequently have considerable overlap with non-coding intronic regions. To define a set of putative well-covered introns, a 10-bp buffer was added upstream and downstream of all gencode v19 coding sequence (to avoid classifying any critical splice sites in the control introns) and this coding sequence was subtracted from the exome probes. Furthermore, any introns within known developmental disorder genes (the DDG2P gene set¹⁷) were excluded. This set of control introns was filtered to include only elements 30 bp in length or larger with >30× coverage.

Evolutionary conservation of non-coding elements. The degree of evolutionary conservation across vertebrates at the element level was calculated using the phastCons vertebrate 100-way score. Scores were retrieved in R using the Bioconductor⁴⁰ package phastCons100way.UCSC.hg19¹⁹.

Benchmarking CADD and other variant scoring methods using MAPS. Scores for all possible SNVs genome-wide were downloaded from CADD²⁴ (<http://cadd.gs.washington.edu/download>), Genomiser⁴¹ (<https://charite.github.io/software-remm-score.html#download>), and fathmm-MKL⁴² (<https://github.com/HASHihab/fathmm-MKL>).

Functional genomic annotation. Data from DNase hypersensitivity assays (broadPeak set, FDR 1%) were downloaded from the Roadmap Epigenome Project³⁰ ftp site (<http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/broadPeak/>) in order to predict regulatory function and tissue specificity in the enhancers and CNEs. The GenomicRanges Bioconductor package was used to intersect DHS peaks with the elements sequenced in this analysis. All code used in this analysis can be found at <https://github.com/pjshort/DDDNonCoding2017>.

Chromatin state predictions (chromHMM 15-state model³⁰) for 111 different tissue types were downloaded from the Roadmap Epigenome Project²⁵ (REP) ftp site (http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/indivModels/default_init/). We considered a CNE to be inactive in a given tissue if it was completely contained within a chromHMM segment described as quiescent, heterochromatin, or polycomb repressed ('9_Het', '13_ReprPC', '14_ReprPCWk', and '15_Quies') in the 15-state model. Using the GenomicRanges⁴³ Bioconductor package and coding sequence from gencode v19, we calculated the distance of each active and broadly inactive element to the nearest exon or transcription start site. All code used in this analysis can be found at <https://github.com/pjshort/DDDNonCoding2017>.

Variant calling, QC, and filtering for unaffected parents. Mapping of short-read sequences was carried out using the Burrows–Wheeler Aligner (BWA; version 0.59) algorithm with the GRCh37 1000 Genomes Project phase 2 reference. The Genome Analysis Toolkit (GATK; version 3.1.1) and SAMtools (version 0.1.19) were used for sample-level BAM improvement. Ensembl Variant Effect Predictor (VEP) based on Ensembl gene build 76 was used to annotate variants and, in coding regions, the transcript with the most severe consequence was selected. We determined the number of variants called per individual, and excluded unaffected parents with variant counts on the extremes of the distribution (top 1% and bottom 1%). We identified a trinucleotide-specific error mode (GTN→GGN) that introduced false positives, which was corrected by strict strand filtering (FS < 20). Across the 7,080 unaffected parents that passed quality control filters, we identified 1,520,250 unique variants in the targeted non-coding elements and coding regions.

MAPS metric within functional annotations. We used DHSs from 39 different tissues to annotate unique variants from 7,080 unaffected parents in the 4,307 CNEs and 595 enhancers we targeted in this analysis as 'in peak' or 'outside peak'. We calculated the MAPS²² for each annotation (within DHS peak in any tissue, outside DHS peak in all tissues, and within DHS peak in each specific tissue).

Where possible, we grouped individual tissues into larger tissue groups based on the Roadmap Epigenome Project²⁵ (REP) categorization. Within the fetal brain tissues (E081, E082 in REP), we calculated the MAPS score for each of the 15 states in order to assess differences in purifying selection between elements that are likely to be inactive versus those that are active.

Defining exome-positive and exome-negative probands. We used the Developmental Disorders Genotype-to-Phenotype Database (DDG2P) to define a set of high-confidence developmental disorder genes (<https://decipher.sanger.ac.uk/ddd#ddgenes>). We classified all probands as 'exome-negative' if they did not have a protein-altering (stop-gain, splice site, or missense variant) DNM in a DDG2P gene with a monoallelic loss of function mechanism, a rare inherited biallelic variant in a DDG2P gene with a recessive mechanism, or a copy number variant identified by clinical microarray and determined to be pathogenic.

De novo mutation calling. *De novo* mutations were called as described¹⁸, excluding variants with posterior probability < 0.00781 as annotated by DeNovoGear⁴⁴.

Trinucleotide germline mutation rate model with CpG-methylation status. A previously described germline mutation rate model based on trinucleotide context²⁶ was adapted to include a correction at CpG sites for methylation status. This method models the null mutation rate at a given site as a Poisson rate parameter that is dependent on the trinucleotide context, where the second base is mutated. We fit a linear model to the ratio of observed/expected variants at MAF < 0.1% in CpG sites based on their methylation status in embryonic stem cells. For all CpG sites, we corrected the trinucleotide mutation rate based on the methylation status to produce a methylation-aware mutation rate model. As the sum of Poisson random variables is Poisson, the rate parameter for a given element, or set of elements, can be determined by summing the mutation rate for each individual site. Simulated mutations were based on the same trinucleotide mutation framework and implemented in an R software package (<https://github.com/pjshort/DenovoSim>).

Testing for enrichment of mutagenic genomic features. The CNEs sequenced in this analysis were intersected with four genomic features previously associated with hypermutability: H3K9me3, H3K27me3, replication timing, and recombination rate. A χ^2 -test was used to test whether elements in which DNMs were observed were enriched for H3K9me3 or H3K27me3 peaks compared to elements in which no DNM was observed using primary mononuclear cells from peripheral blood (Roadmap Epigenome ID E062). For replication timing²⁸ and recombination rate²⁹, a Wilcoxon rank sum test was used to test for differences between the two sets of elements.

Testing for hypermutability using rare variation in deep whole genomes. We calculated the number of observed rare variants (MAF < 0.1%) per unit of expected mutability from the null mutation model for the fetal brain-active and brain-inactive elements that we sequenced in 7,509 non-Finnish European deep whole genomes present in the gnomAD data set. We used bootstrap re-sampling to estimate the standard error around the estimated rare variants per unit mutability for each set. To assess the power of this approach to detect mutability, we simulated rare variants using the null mutation model under 1.1×, 1.2×, and 1.3× mutability in the fetal brain-active elements and tested the power to reject the null hypothesis of mutability 1.0× for different numbers of elements from 50 to 1,000 in steps of 50.

Statistical testing for mutational burden. The *P* value for the number of observed *de novo* mutations compared to expected is calculated in R as: $\text{ppois}(n_{\text{obs}} - 1, \text{lambda} = \mu, \text{lower.tail} = \text{FALSE})$ where n_{obs} is the number of observed mutations within an element and μ is the mutability of the element(s) being tested (under the null model described above) multiplied by the number of probands. The burden testing we performed across subsets of elements and phenotypes included multiple nested hypotheses that were accounted for with a conservative Bonferroni-adjusted *P* value threshold based on the number of explicit and implicit tests. We corrected for thirteen tests within the exome-negative cohort based on branching on element class and phenotype, where appropriate (detailed in Supplementary Fig. 5). In testing for single elements with an excess of observed mutations, we employed a conservative Bonferroni adjustment to correct for 2,613 tests (the number of fetal brain-active CNEs).

Defining fetal brain-active elements. We used the Roadmap Epigenome Project²⁵ DNase data and chromHMM annotations to annotate the CNEs as 'active' and 'inactive' in the fetal brain. We defined all of the sections of the genome predicted to be quiescent, heterochromatin, or polycomb repressed ('9_Het', '13_ReprPC', '14_ReprPCWk', and '15_Quies' in the 15-state model) as 'inactive' states. We considered a CNE or enhancer to be inactive if it was completely contained within an inactive chromHMM³⁰ segment in both male and female fetal brain and if it did not overlap with any high confidence DNase hypersensitive site in male or female fetal brain. In total, 2,613 of 4,307 CNEs and 383 of 595 experimentally validated enhancers were predicted to be active in the fetal brain based on these criteria. All code used in this analysis can be found at <https://github.com/pjshort/DDDNonCoding2017>.

Estimating the proportion of fetal brain-active CNEs acting as enhancers.

To evaluate the proportion of CNEs that may be acting as enhancers, we analysed downloaded enhancer RNA (eRNA) data from the fetal brain generated by the FANTOM5 consortium³² and predicted fetal brain enhancers from EnhancerAtlas³³, which combines multiple sources of data to identify enhancers in different tissues. We used the experimentally validated fetal brain-active VISTA enhancers to estimate this sensitivity of each data set. We then overlapped the fetal brain-active CNEs with the FANTOM5 eRNA and EnhancerAtlas predictions and used the sensitivity estimates to estimate the total proportion of fetal brain CNEs likely acting as enhancers.

Stratifying enhancers by evolutionary conservation. Phastcons vertebrate 100-way scores¹⁹ were retrieved in R using the Bioconductor package phastCons100way.UCSC.hg19 for each of the 383 fetal brain-active experimentally validated enhancers.

Statistical testing for enrichment of recurrently mutated elements. We defined any element observed with a DNM in at least two unrelated probands as 'recurrently mutated'. We used the simulation framework described above to calculate the likelihood of observing a given number of recurrently mutated elements. To calculate the significance of individual elements, we calculated the likelihood of observing n DNMs in a given element with mutability λ in R as $\text{ppois}(n_obs - 1, \lambda = \mu, \text{lower.tail} = \text{FALSE})$. The P values were compared to a genome-wide significance cutoff of 0.05/2,613 or $P < 1.91 \times 10^{-5}$ (Bonferroni-corrected P value based on independent tests for enrichment across 2,613 elements).

Defining CNE clusters. In order to identify clusters of CNEs, we compared the inter-element distance in our set of sequenced CNEs to the inter-element distance of the same number of elements randomly distributed genome-wide. We used agglomerative hierarchical clustering with single linkage clustering in R to define clusters at a given inter-element distance. The false discovery rate (FDR) for a set of clusters can be determined by comparing the number of observed clusters to the number expected under the randomly distributed null model at the same inter-element distance. For this analysis, we used a maximum inter-element distance of 10kb, which corresponds to a false discovery rate of 10%.

chromHMM state of recurrently mutated elements. We used the chromHMM 15-state model predictions from the Roadmap Epigenome Project²⁵ (REP) fetal brain male and female (E081, E082 in REP) to classify each of the DNMs observed in recurrently mutated elements. The predicted state in male/female fetal brain was not always concordant. When one annotation predicted the element as inactive and the other as active, we kept the active prediction. When the DNM was predicted to be active in both male and female fetal brain, but in different states, we chose the male state. Re-running this analysis to instead choose the female prediction did not substantially change the outcome.

Phenotypic similarity by human phenotype ontology comparison. Referring clinicians used the Human Phenotype Ontology (HPO) version 2013-11-30 to systematically describe patients upon recruitment to the DDD study. In order to compare phenotypic similarity between groups of patients statistically, the hpo similarity test was used⁴⁵.

Clustering of DNMs. To test the observed DNMs for clustering that might imply disruption of an underlying binding site or functional motif, we used the denovonear framework described previously¹⁸. This method compares the distance between observed DNMs to the distance between simulated DNMs based on the trinucleotide null mutation model to generate an empirical P value.

Gene target prediction and pair-wise overlap. We used four different methods of gene target prediction to link CNEs and enhancers to putative target genes.

The first method, Genomicus, predicts gene targets based on evolutionary conservation with nearby genes. Genomicus determines the extent to which each CNE is within the same syntenic block with nearby genes across a number of vertebrate species and predicts one or more targets¹⁴. The Genomicus method produces at least one prediction for 90% of CNEs (approximately one-third of these are the closest genes).

The second method compares DNase hypersensitivity at each CNE to expression of nearby genes in 56 different tissues (using RNA sequencing (RNA-seq)) to search for CNE-gene pairs that show a correlation between DNase signal and gene expression³⁴. This method produces statistically significant predictions for only 28% of CNEs in our set and is likely to be underpowered to detect elements that are active in specific tissues or time points.

The third method is to link CNEs to putative target genes using chromatin interaction data (Hi-C) in two different regions of the fetal brain⁴⁶. The use of Hi-C data is the most direct and tissue-specific of all of the prediction methods used, but the prediction is sparse (26% of CNEs with evidence of fetal brain activity have a predicted target).

The fourth method used is a simple heuristic to choose the gene with the closest TSS (for intergenic elements) or the gene containing the element

(for introns). Choosing the closest gene allows us to make a prediction for 100% of elements, but comparison with chromatin conformation and DHS-based methods has shown that the closest gene is likely to be the target in 7% and 12% of cases, respectively^{34,47}.

We used the Genomicus, DHS, and Hi-C predictions to generate aggregated predictions which we considered 'high confidence' if predicted by at least two of the three methods.

To assess the pair-wise concordance reported in Extended Data Fig. 7, we took the set of CNEs for which at least one gene target was reported in both methods and tested how frequently both methods identified the same gene within the set of predicted targets.

Brain developmental expression trajectory. BrainSpan developmental RNA-seq data (<http://www.brainspan.org>) were processed as previously described³⁵. Expression values were log-transformed ($\log_2[\text{RPKM}+1]$) and scale normalized. This expression data set consists of six brain regions (cortex, thalamus, striatum, hippocampus, amygdala, and cerebellum) and developmental epochs that span prenatal (8–37 post-conception week) and postnatal (4 months–40 years) periods. Genes that are associated with CNEs with DNMs and without DNMs were selected and their developmental expression trajectories were plotted using loess smooth. **Transcription factor binding analyses.** The JASPAR2016 and TFBSTools Bioconductor packages⁴⁸ were used to retrieve position weight matrices for 454 human transcription factors. Analyses in this paper focus on the 202 transcription factors predicted to be expressed in the brain (cortex-expressed from GTEx data set⁴⁹).

A custom R package called 'denovoTF' (<https://github.com/pjshort/denovoTF>) was written to predict any change in transcription factor binding at sites where DNMs were observed or simulated. This analysis works by scanning the reference and alternative sequences for all 202 PWMs and comparing predicted binding events on both sequences. By comparing the potential binding affinity for ref and alt sequences, we can predict loss of binding (alt binding < ref binding), gain of binding (alt binding > ref binding), and silent (no difference). 'Silent' DNMs fall into two classes: those for which binding is predicted on both reference and alternate, but strength of binding is unchanged, and those which do not lie in a predicted transcription factor binding site.

The analysis of motif enrichment (AME) tool from the meme suite was used to identify a subset of PWMs that was significantly enriched in fetal brain-active elements⁵⁰. Comparing the fetal brain-active CNEs to the fetal brain inactive CNEs returned a set of 90 transcription factors, of which 45 were expressed in the brain and had PWMs available in JASPAR2016⁴⁸. This analysis was performed on the meme-suite web server using the following command:

```
ame-verbose 1-oc.-control meme_chromHMM_fb_inactive_all.fasta-bgformat 1-scoring avg-method ranksum-pvalue-report-threshold 0.05 meme_chromHMM_fb_active_all.fasta db/JASPAR/JASPAR_CORE_2016.meme
```

In order to test for enrichment of loss of binding or gain of binding events in the observed DNMs, we compared predicted impact on transcription factor binding in observed DNMs to 1,000 simulations of mutations across the 2,613 fetal brain-active elements for 6,147 probands.

Nucleotide-level conservation (PhyloP). PhyloP scores represent the $-\log_{10}$ P value that a given site is evolving neutrally³¹. We used a tabix file of pre-computed PhyloP vertebrate 100-way scores for every site in the genome in order to annotate the DNMs observed in exome-negative probands to exome-positive probands as well as the simulated null model.

Power calculations at different study sizes. We used the trinucleotide null model described previously in order to estimate our power to detect disease-associated elements. Parameters that affect power include the fold enrichment for disease-causing mutations in the DDD cohort (proportional to the incidence of severe developmental disorders with a genetic basis in the population), the proportion of mutations within a true disease-associated element expected to be pathogenic, the penetrance of such mutations, the size and mutability of the elements tested, and the number of trios analysed. To estimate the power across different study sizes, we fixed the remaining parameters as follows: 120-fold enrichment for disease-causing mutations, proportion of mutations expected to be pathogenic at 8% (lower bound estimate for coding regions), penetrance at 100%, and the elements tested were the 2,613 fetal brain-active CNEs. Code for power analysis can be found in the R script: https://github.com/pjshort/DDDNonCoding2017/blob/master/analysis_notebooks/Figure4_maximum_likelihood_and_genome_estimate.Rmd.

Likelihood of power calculation model parameters under observed data. To test the likelihood of different models of dominant disease mechanism within the non-coding space we adapted the power calculation framework described above to test the probability of observing our data across two different parameters: the number of elements (out of 2,613) with a dominant disease mechanism and the proportion of mutations expected to be pathogenic. We tested the likelihood

of observing 286 DNMs, 25 recurrently mutated elements, and zero elements at genome-wide significance while systematically varying two parameters: the proportion of mutations expected to be pathogenic parameter (from 0.01% to 10.0% in increments of 0.01%) and the proportion of elements with true disease associations (from 0 to 2,613 in increments of 5). In this analysis, the remaining parameters were held constant: $120\times$ enrichment for pathogenic mutations, penetrance at 100%, testing 2,613 fetal brain-active CNEs, and number of trios at 6,147. Code can be found in the R notebook: https://github.com/pjshort/DDDNonCoding2017/blob/master/analysis_notebooks/Figure4_maximum_likelihood_and_genome_estimate.Rmd.

Estimating the genome-wide burden of DNMs in fetal brain-active elements.

First, we intersected all targeted non-coding sequences, irrespective of original class, with fetal brain DHS peaks. We used the phastcons100 score (scores retrieved in R using the Bioconductor package phastCons100way.UCSC.hg19¹⁹) to rank these elements by evolutionary conservation. The ratio of observed/expected DNMs was computed with a sliding window across the elements (window size of 1,000 elements, shift of 100 elements). This approach resulted in a median of 62 DNMs expected in each bin (minimum 51, maximum 68) which was compared to the observed number of DNMs. We fit a logistic regression to the excess observed/expected in each window, setting any window with observed less than expected to have an excess of zero. We used the logistic regression fit on the CNEs sequenced in our analysis to predict the burden of DNMs in this genome-wide set.

Transmission of rare variants. All variants that were heterozygous in one parent were tested for any patterns of overtransmission within different variant classes. Only elements with $>20\times$ median coverage were used for this analysis, as elements without adequate coverage showed systematic underestimation of transmission. The observed proportion of rare variants that were transmitted from parents to affected probands was compared to the expected proportion under the null hypothesis (50%) using a binomial test.

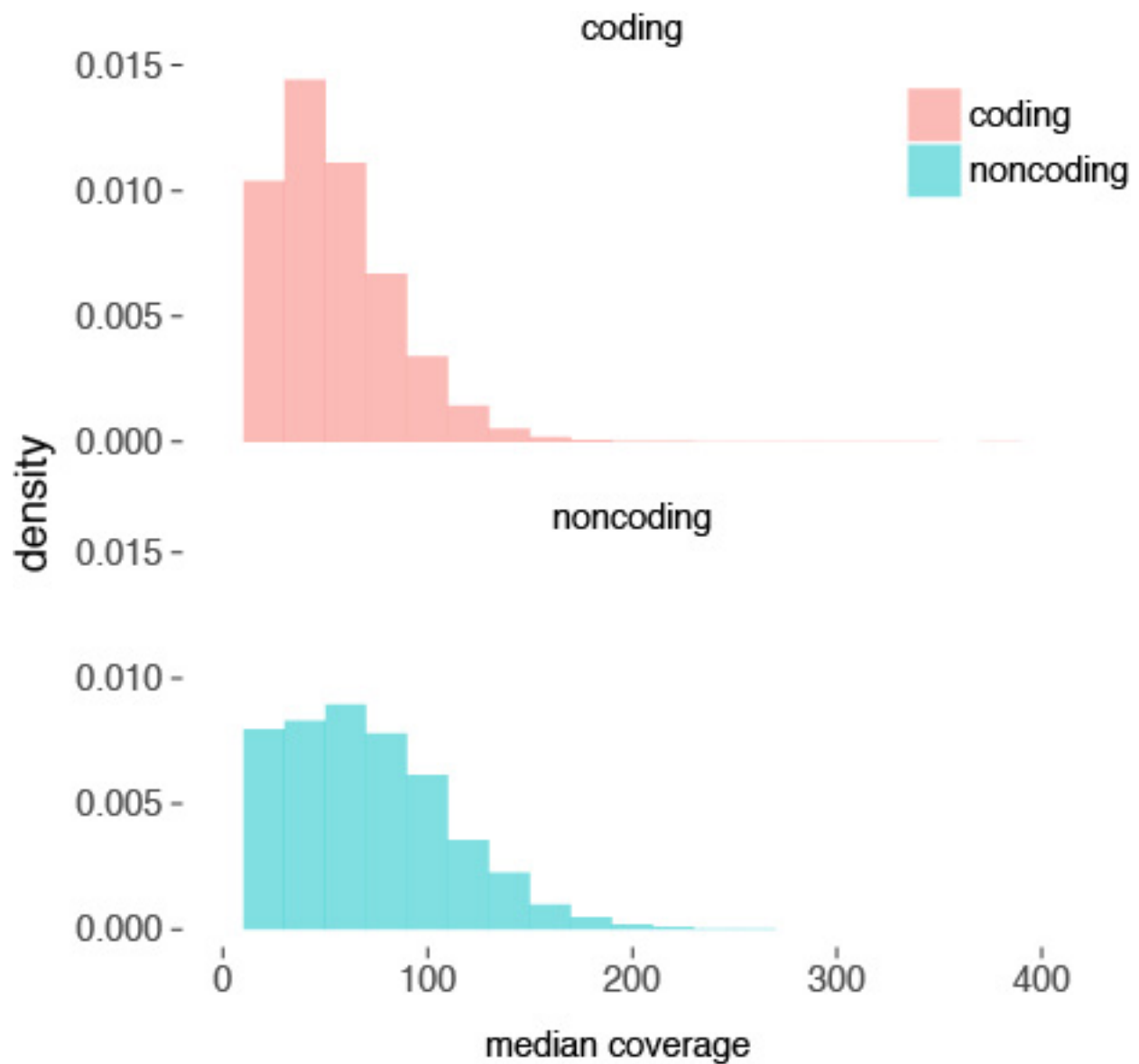
Testing CNEs for 'already perturbed haplotypes'. In order to test the hypothesis that DNMs in fetal brain-active CNEs may be contributing to a developmental disorder via a second hit on an already weakened haplotype, we extracted the rare variants present in the relevant DNM-containing CNE for each proband. We compared the proportion of probands with at least one variant besides the observed DNM in the fetal brain-active CNEs compared to the fetal brain-inactive CNEs. We also calculated the total burden of rare variation within the DNM-containing

element (measured as SNVs per kb) for probands with DNMs in fetal brain-active CNEs compared to probands with DNMs in fetal brain-inactive CNEs.

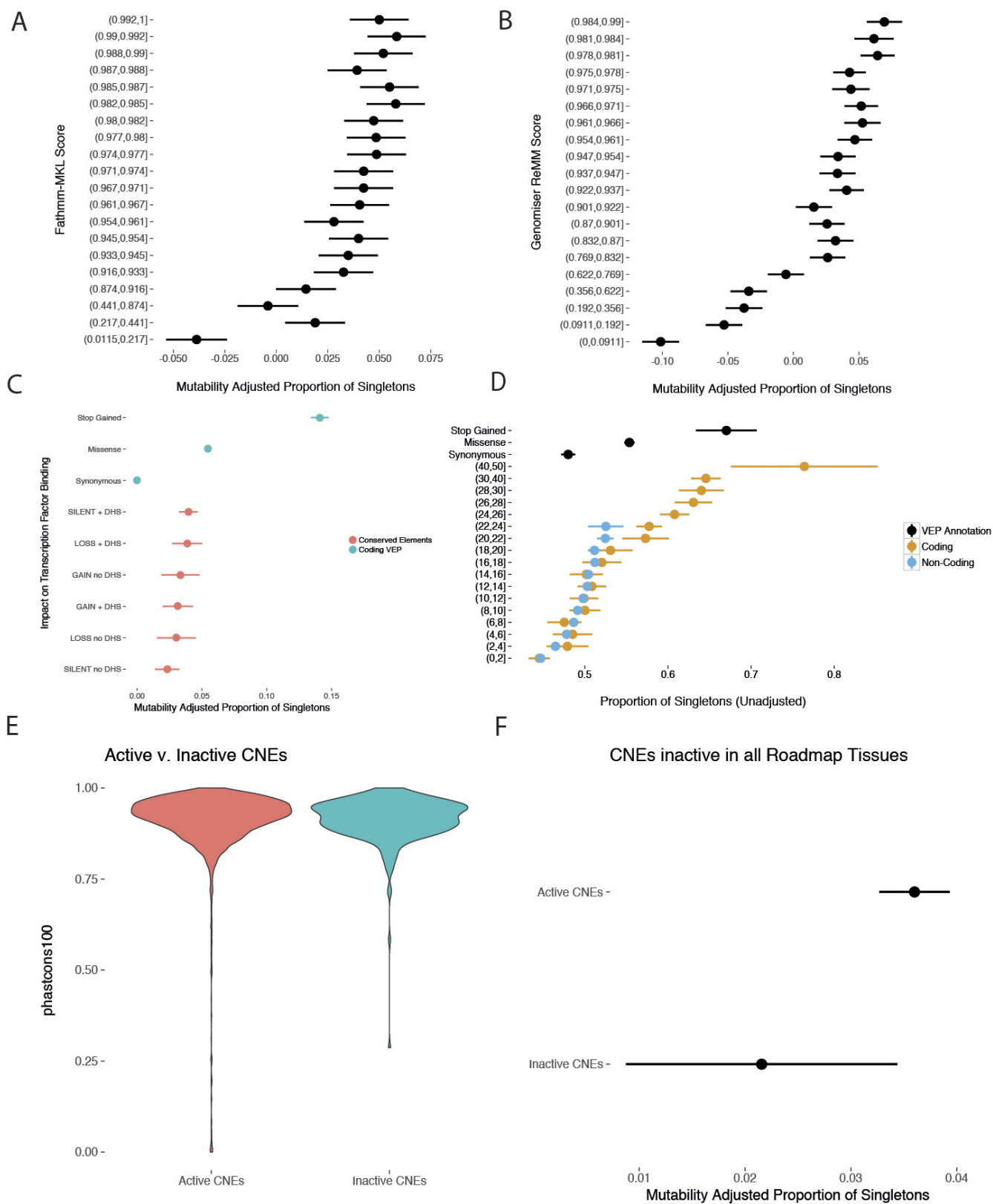
Code availability. Source code used to analyse data and generate the figures for this article can be found at <https://github.com/pjshort/DDDNonCoding2017/>.

Data availability. Sequencing and phenotype data are accessible via the European Genome-phenome Archive (EGA) under study number EGAS00001000775 (<https://www.ebi.ac.uk/ega/studies/EGAS00001000775>). The DDG2P gene list of genes associated with developmental disorders is available at www.ebi.ac.uk/gene2phenotype. All other data are available from the corresponding author upon request.

39. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
40. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
41. Smedley, D. *et al.* A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am. J. Hum. Genet.* **99**, 595–606 (2016).
42. Shihab, H. A. *et al.* An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536–1543 (2015).
43. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
44. Ramu, A. *et al.* DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat. Methods* **10**, 3–7 (2013).
45. Akawi, N. *et al.* Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat. Genet.* **47**, 1363–1369 (2015).
46. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
47. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
48. Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44** (D1), D110–D115 (2016).
49. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
50. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).

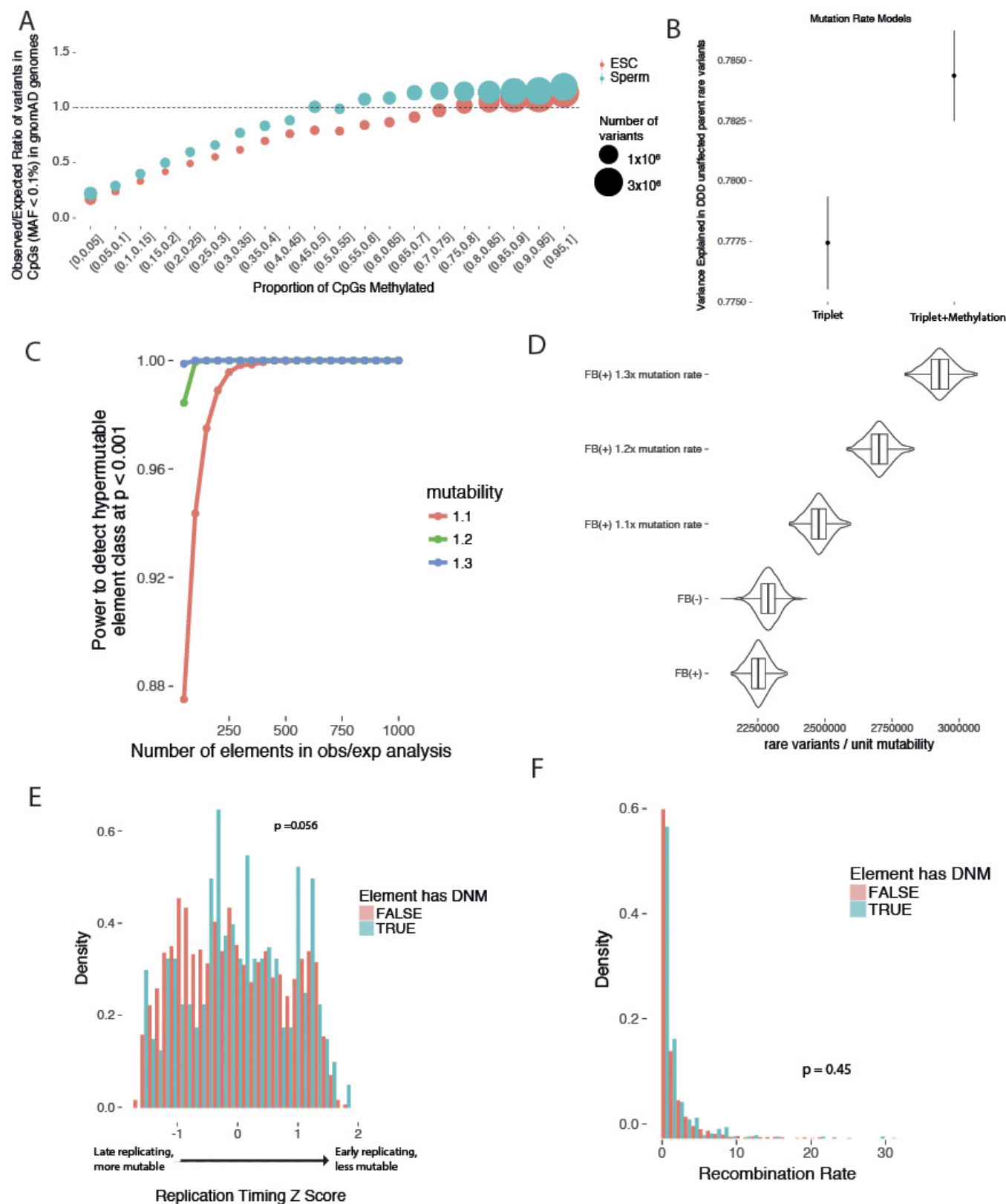


Extended Data Figure 1 | Coverage in targeted non-coding elements. Coverage in the targeted non-coding elements is comparable to the protein-coding exons (median 73× and 56×, respectively).



Extended Data Figure 2 | Assessment of variant deleteriousness metrics and selective pressure in CNEs. Dots and bars represent the point estimate and 95% CI, respectively, for MAPS and proportion singletons. **a, b**, Fathmm-MKL (**a**) and Genomiser (**b**) separate benign variation (low MAPS score) from likely damaging variation (high MAPS score), but do not identify any classes of variation under strong selective constraint. **c**, There was no significant difference in the strength of purifying selection measured by MAPS between sites predicted to result in loss, gain, or no change in transcription factor binding. **d**, Validation of Fig. 1c using

whole-genome data from the UK10K project. While CADD can identify coding variation under strong selective constraint (as measured by the proportion of singletons), CADD is unable to identify strongly constrained non-coding variants. **e, f**, The subset of CNEs sequenced in the DDD cohort that are predicted to be inactive in all 111 Roadmap Tissues ($n = 261$) exhibit a similar degree of evolutionary conservation (**e**) but lower selective constraint (**f**) in a healthy population compared to CNEs active in at least one tissue ($n = 4,046$).

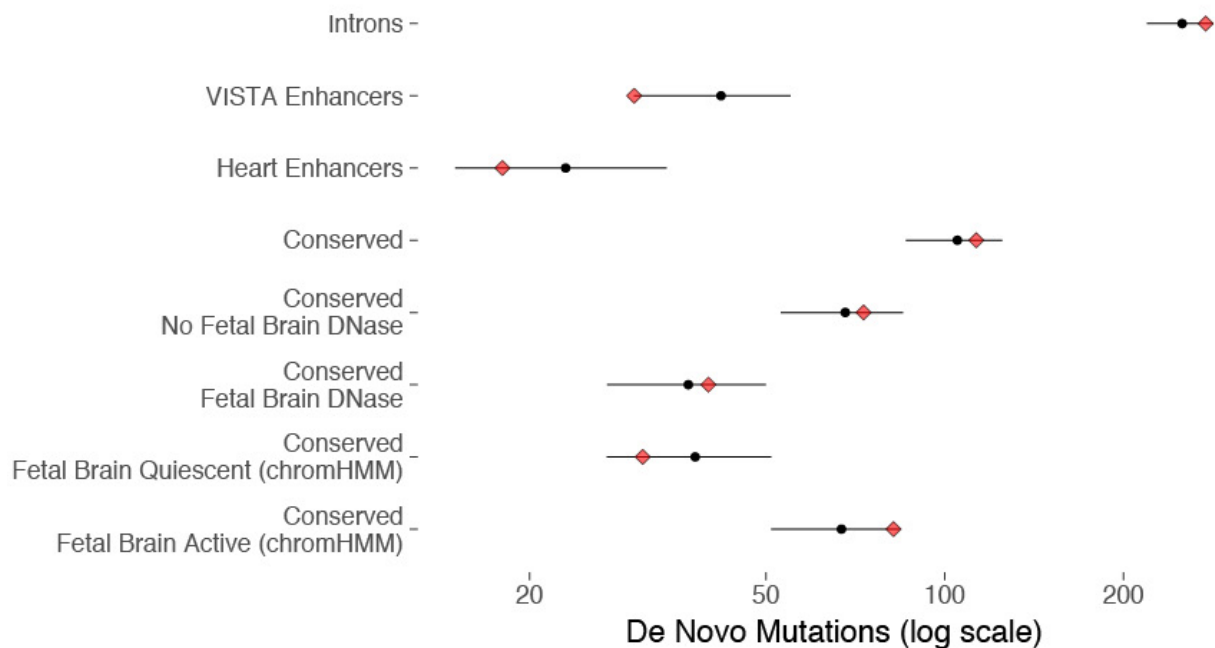


Extended Data Figure 3 | Genomic factors that affect mutation rate in non-coding elements. **a**, Aggregating CpG sites genome-wide into bins of methylation proportion from 0% (unmethylated in all cells) to 100% (methylated in all cells) and calculating the observed/expected ratio reveals differences in mutability not accounted for by a triplet model alone. **b**, A mutation rate model incorporating a correction for CpG methylation explains greater variance in rare variant counts in the DDD unaffected parents. **c**, Levels of rare variation in deep whole genomes ($n = 7,509$ non-Finnish Europeans) were used to estimate power to detect

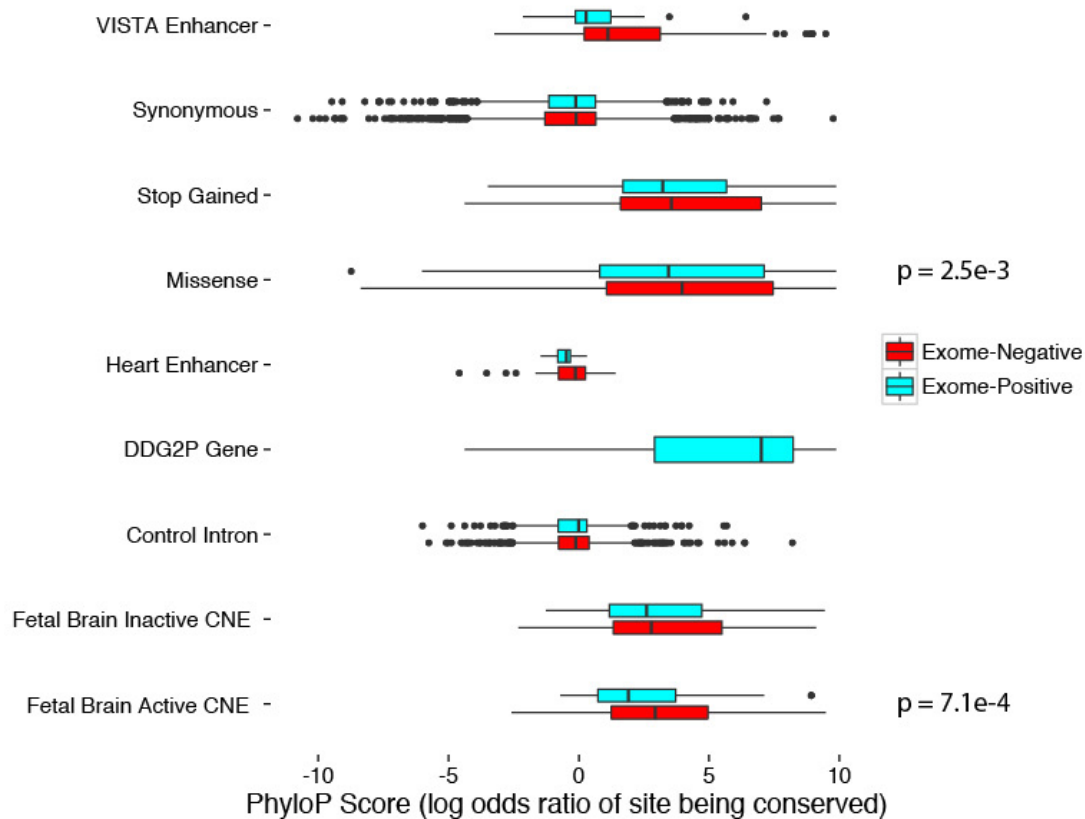
a hypermutability of 1.1 \times , 1.2 \times , or 1.3 \times . **d**, The level of rare variation in the fetal brain-active elements ($n = 2,613$, FB(+)) is slightly lower than in the fetal brain-inactive elements ($n = 1,694$, FB(-)), consistent with similar mutability between the two element sets with slightly stronger purifying selection in the fetal brain-active elements. **e**, **f**, Elements with DNMs observed in our study are not enriched in late-replicating regions (**e**) or in regions with higher recombination rate (**f**), which have been shown to be hypermutable.

A

De Novo Mutations in Non-Coding Elements

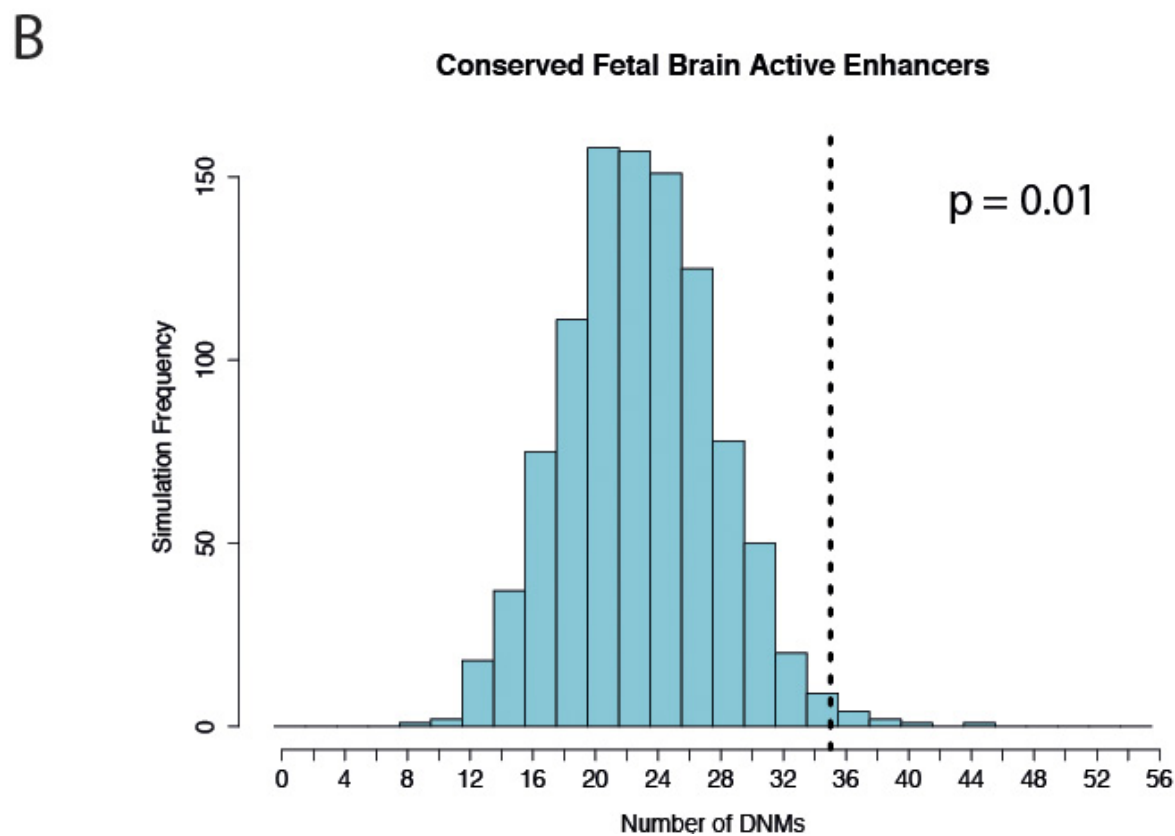
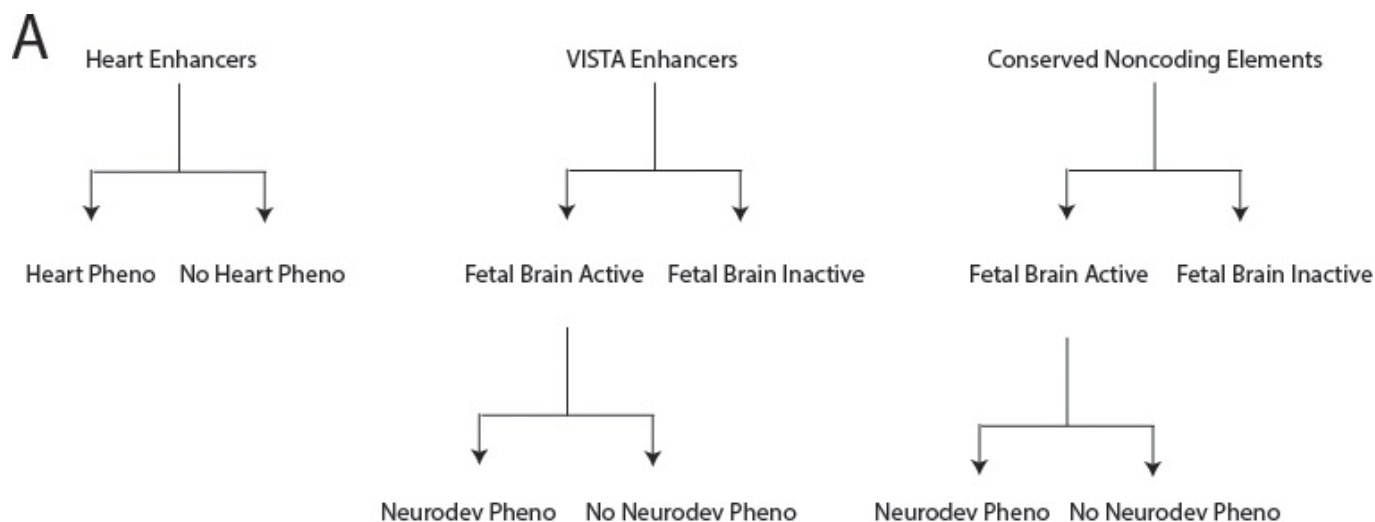


B



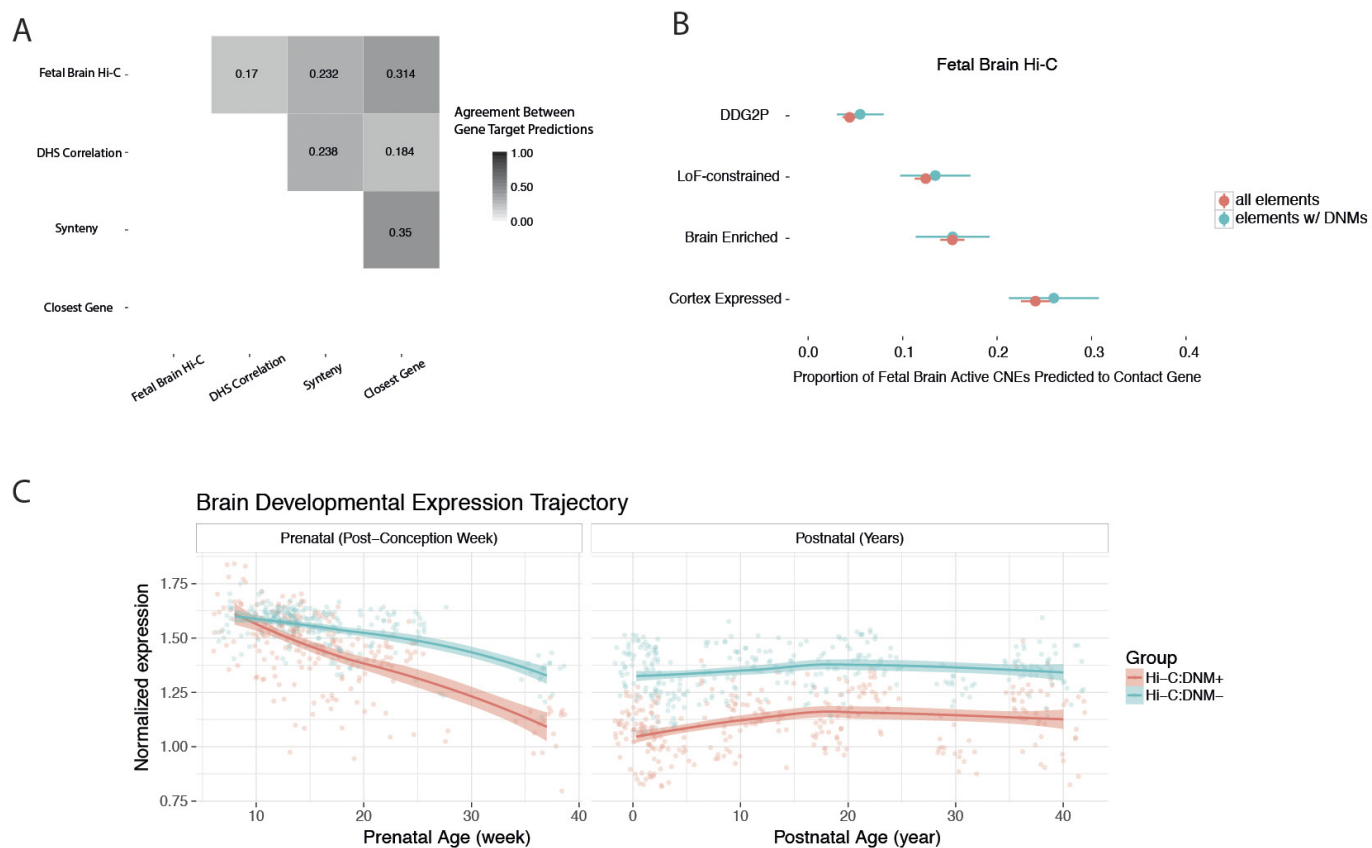
Extended Data Figure 4 | Non-coding mutations in exome-positive probands and poorly evolutionarily conserved sites make a minimal contribution to severe developmental disorders. a, In the 1,691 'exome-positive' probands, there is no evidence for a burden of DNMs in any of the non-coding element classes tested. Red diamonds indicate the observed counts, while black circles and bars indicate the expected count and 95%

CI, respectively. b, DNMs in exome-negative probands show a greater degree of evolutionary conservation (measured by PhyloP score) than DNMs in exome-positive probands in two classes: fetal brain-active CNEs (median 1.57 exome-positive, 2.85 exome-negative, $n = 368$ mutations) and missense changes (median 3.43 exome-positive, 3.98 exome-negative, $n = 6,244$ mutations).



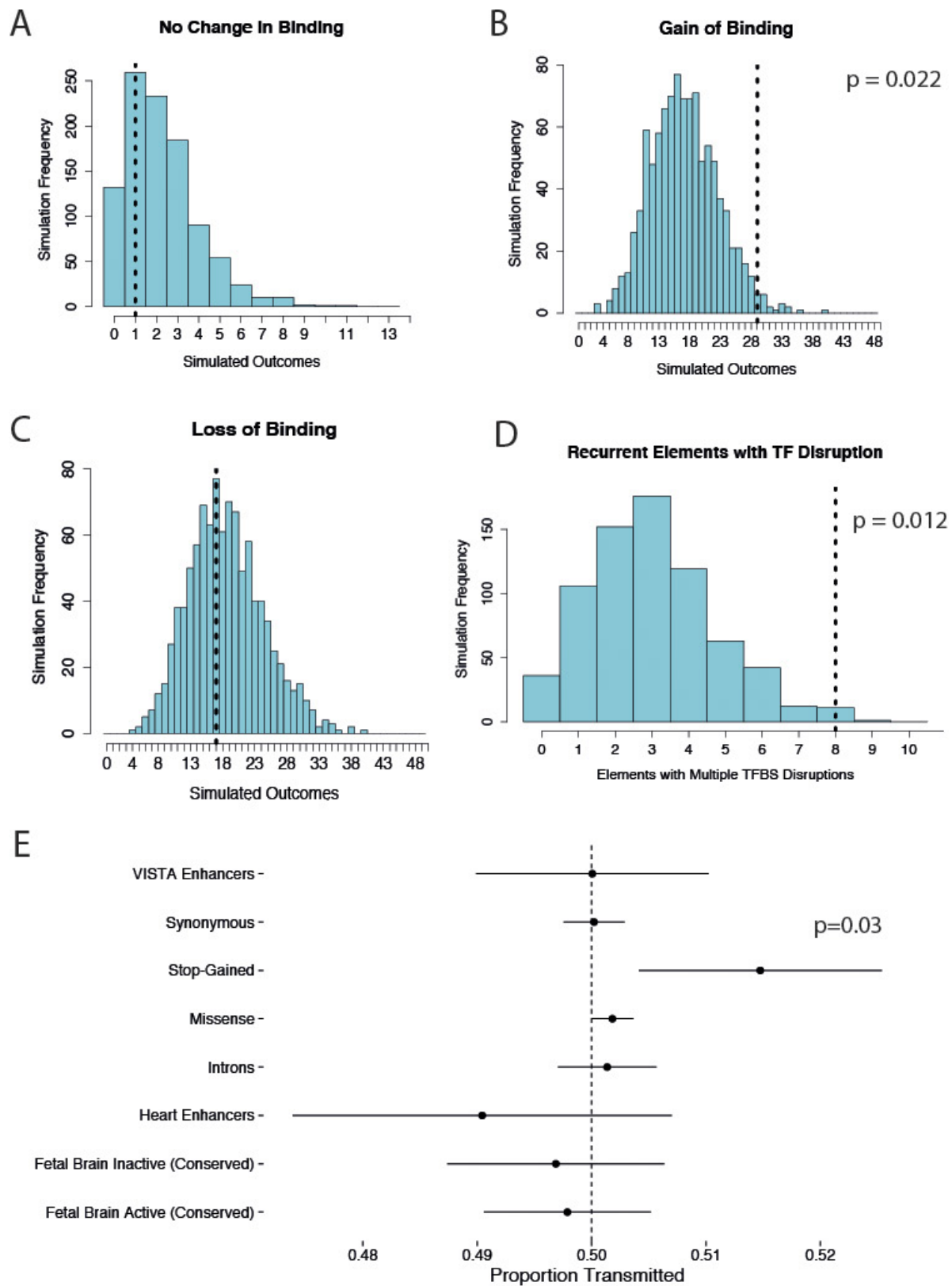
Extended Data Figure 5 | Hypothesis test enumeration and enrichment for mutations in highly conserved fetal brain-active enhancers. a, We corrected for thirteen tests in order to account for the nested hypotheses

based on element class and phenotype in this analysis. **b,** Evolutionarily conserved fetal brain-active enhancers ($n = 106$) are enriched for DNMs in exome-negative probands.



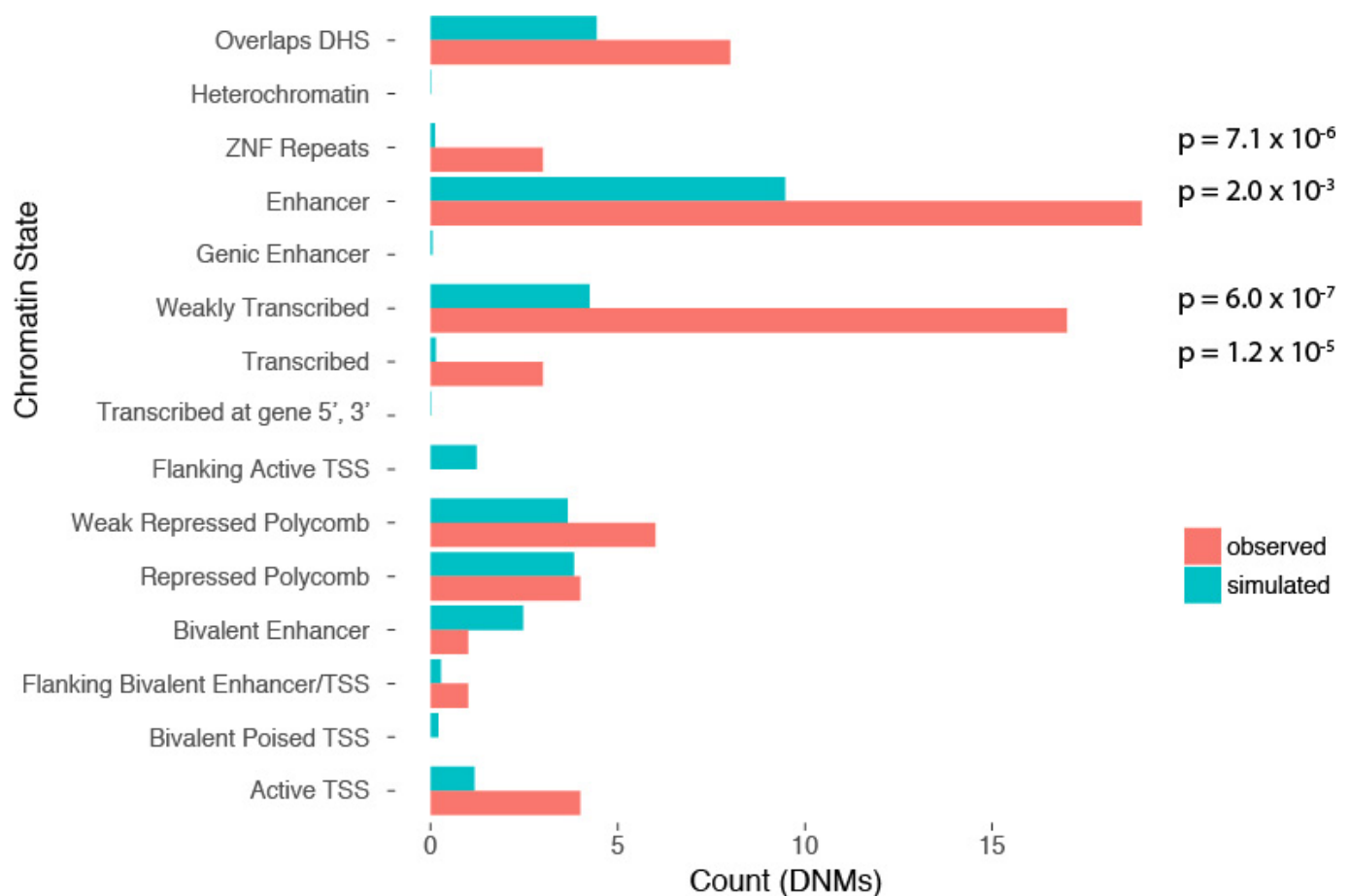
Extended Data Figure 6 | Gene target prediction for targeted non-coding elements. Pairwise concordance between four different gene target prediction methods is low. Using predicted targets from fetal brain Hi-C data, elements with an observed DNM in exome-negative probands

($n = 286$) do not show any bias towards any of the gene sets consistently implicated in neurodevelopmental disorders. Dots and bars represent the point estimate and 95% confidence interval, respectively.



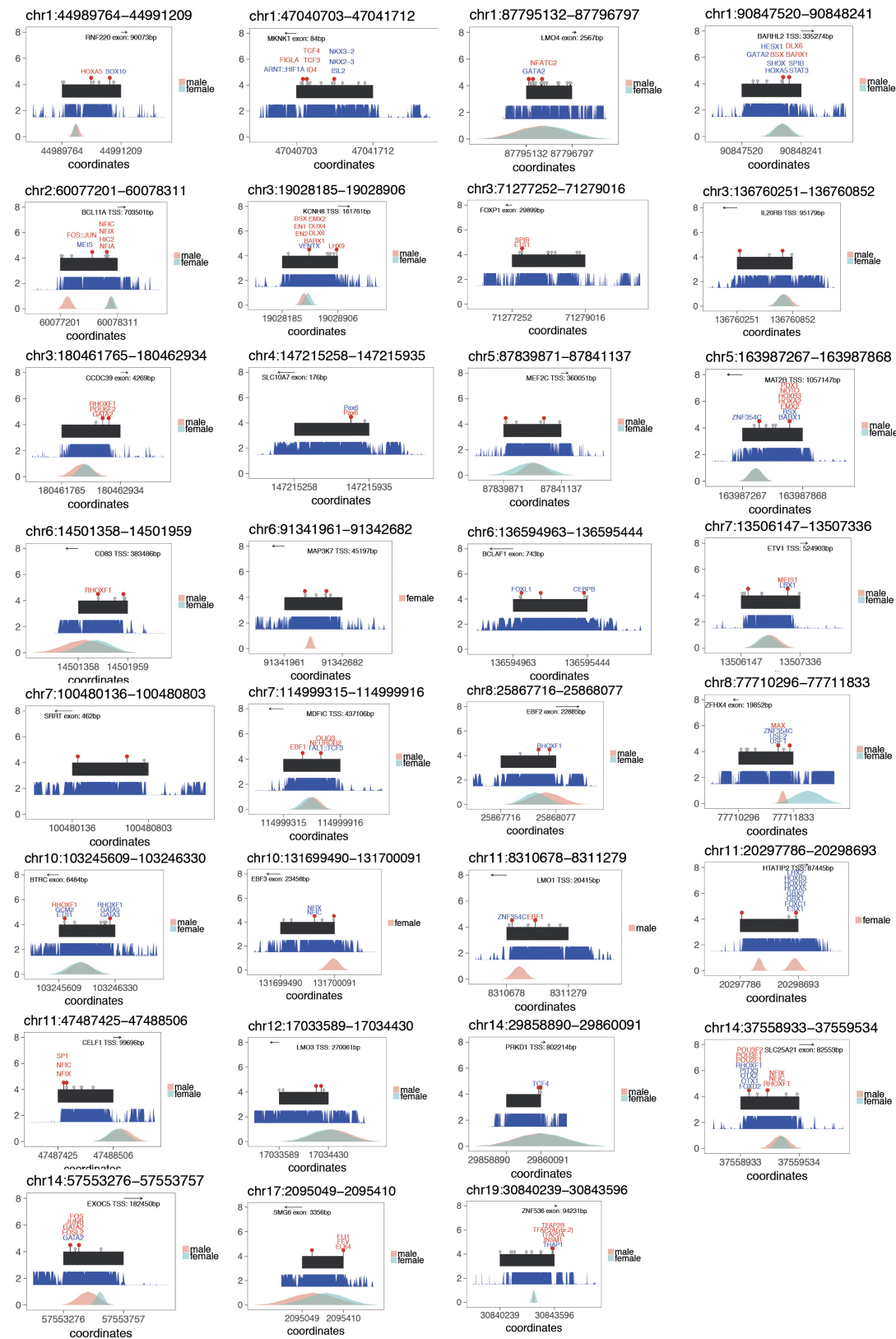
Extended Data Figure 7 | Transcription factor binding disruption and transmission disequilibrium test. a–d, Comparison of predicted change in transcription factor binding for observed DNMs compared to null mutation model. Empirical P values derived from comparison with mutations simulated from the null mutation model. e, None of the non-

coding element classes tested show any evidence of overtransmission from parents to affected children. Dots and bars represent the point estimate and 95% confidence intervals of estimates of transmission proportions, respectively.



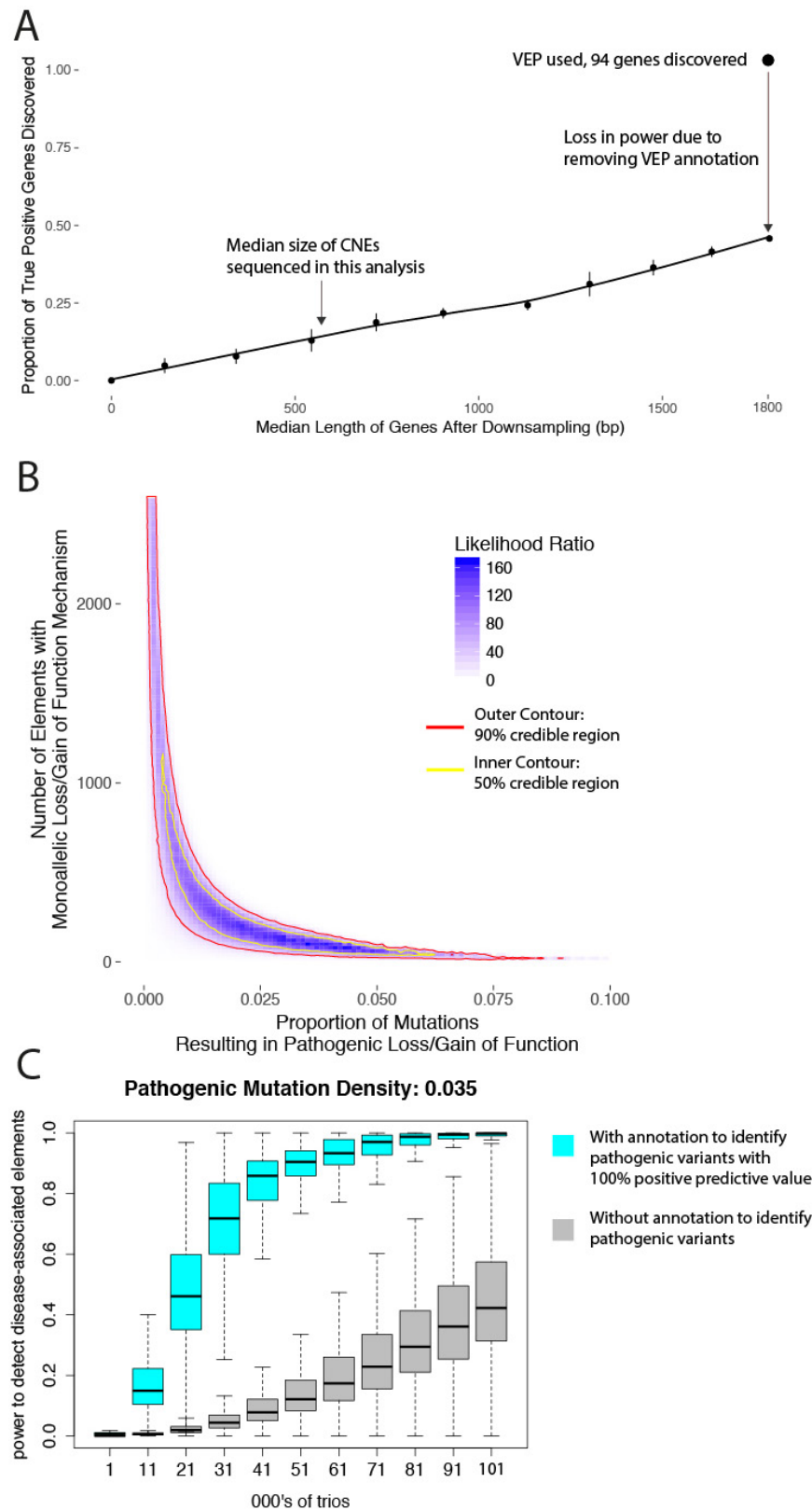
Extended Data Figure 8 | Predicted chromatin state for recurrently mutated elements. chromHMM state of the $n = 31$ recurrently mutated elements shows enrichment for enhancers and transcribed elements. Elements that overlapped a high confidence DHS but were predicted as

quiescent by chromHMM are classed as Overlaps DHS. P values derived from Poisson distribution with parameter lambda defined by the simulated data.



Extended Data Figure 9 | Schematic describing each of the thirty-one recurrently mutated elements. Element is in black, red lollipops denote observed DNMs, grey lollipops denote observed variation at MAF > 0.1%

in 7,080 unaffected parents, phastcons100 conservation score is shown in blue, and DHSs from the Roadmap Epigenome project are shown in blue/pink in the bottom track.



Extended Data Figure 10 | Empirical and simulated power for disease association in targeted non-coding elements. a, Estimation of the reduction in power due to size differences between non-coding elements and genes (median 600 bp versus 1,800 bp) and ignoring VEP annotations used to stratify benign from likely damaging variants. Dots and bars represent the point estimate and 95% confidence interval, respectively. **b,** Credible intervals for the proportion of fetal brain-active conserved

elements and proportion of sites within those elements with a dominant mechanism for developmental disorders. **c,** Power calculations for disease-associated non-coding element discovery. Without annotation or tools to discriminate pathogenic from benign variants in non-coding elements (grey), more than 100,000 trios are required to achieve 40% power. With annotation or tools to fully discriminate likely pathogenic from benign variants (blue), 40% power is achieved with only 21,000 trios.

Encoding of danger by parabrachial CGRP neurons

Carlos A. Campos¹, Anna J. Bowen¹, Carolyn W. Roman¹ & Richard D. Palmiter¹

Animals must respond to various threats to survive. Neurons that express calcitonin gene-related peptide in the parabrachial nucleus (CGRP^{PBN} neurons) relay sensory signals that contribute to satiation and pain-induced fear behaviour, but it is unclear how they encode these distinct processes. Here, by recording calcium transients *in vivo* from individual neurons in mice, we show that most CGRP^{PBN} neurons are activated by noxious cutaneous (shock, heat, itch) and visceral stimuli (lipopolysaccharide). The same neurons are inhibited during feeding, but become activated during satiation, consistent with evidence that CGRP^{PBN} neurons prevent overeating. CGRP^{PBN} neurons are also activated during consumption of novel foods or by an auditory cue that has previously been paired with electrical footshocks. Correspondingly, silencing of CGRP^{PBN} neurons attenuates the expression of food neophobia and conditioned fear responses. Therefore, in addition to transducing primary sensory danger signals, CGRP^{PBN} neurons promote affective-behavioural states that limit harm in response to potential threats.

The parabrachial nucleus (PBN) receives ascending inputs from many sensory modalities and relays this information to brain areas that transform sensory signals into autonomic, affective, and behavioural responses^{1,2}. We have focused on calcitonin gene-related peptide (CGRP)-expressing neurons in the external lateral PBN, initially because disinhibition of these neurons following ablation of hypothalamic neurons that express orexigenic agouti-related peptide (AgRP) causes severe anorexia³. Subsequently, we found that activation of CGRP^{PBN} neurons contributes to meal termination⁴, indicating that these neurons play a physiological role in controlling appetite. Additional studies have shown that CGRP^{PBN} neurons are directly activated by neurons that reside in the nucleus of the solitary tract (NTS)⁵, which receives visceral satiety signals via vagal sensory afferents⁶. Notably, CGRP^{PBN} neurons are also part of a spino-parabrachial pathway that is critical for eliciting fear responses to cutaneous pain⁷. CGRP^{PBN} neurons have axonal projections that activate the central amygdala and the bed nucleus of the stria terminalis (BNST)³, forebrain nuclei that have been implicated in both feeding and pain-related sensory processing^{7–9}. These observations raised the question of whether subpopulations of CGRP^{PBN} neurons relay specific stimuli to the forebrain.

We hypothesized that CGRP^{PBN} neurons perform a common function of signalling danger, whether it relates to cutaneous pain or consumption of food beyond physiological capacity, and reasoned that visualization of the calcium dynamics (correlates of neural activity) of individual neurons *in vivo*^{10,11} would allow us to investigate whether subpopulations of CGRP^{PBN} neurons respond to specific stimuli. We show that the same neurons respond to feeding and various noxious or painful stimuli. CGRP^{PBN} neurons also respond to danger signals that require learning and memory. Together with loss-of-function studies using transgenic and viral techniques, these results show that activation of CGRP^{PBN} neurons contributes to affective-behavioural states that limit harm.

Coding of painful stimuli by CGRP^{PBN} neurons

We injected adeno-associated virus (AAV) expressing Cre-dependent GCaMP6m into the lateral PBN of transgenic mice with Cre

recombinase targeted to the CGRP-encoding gene (*Calca*^{Cre/+}), thereby introducing a calcium-sensitive fluorophore specifically into CGRP^{PBN} neurons (Fig. 1a). This strategy allowed us to record calcium dynamics from individual neurons in mice implanted with a graded-index lens (Fig. 1b, c). A tail pinch transiently increased calcium activity in CGRP^{PBN} neurons (Fig. 1d and Supplementary Video 1); all 384 neurons tested responded to at least four out of five tail pinches during a single recording session. Most neurons exhibited increased calcium activity following a pinch to the tail or any of the four paws (Fig. 1e, f). To determine whether CGRP^{PBN} neurons are activated by noxious stimuli on the face, we placed a warm metal rod on the lip; this activated the same neurons that responded to tail pinch (Extended Data Fig. 1a, b). Intraperitoneal injection of lipopolysaccharide (LPS), which mimics bacterial infection and causes visceral malaise, also activated pinch-responsive neurons (Fig. 1g, h and Extended Data Fig. 1c). Thus, individual CGRP^{PBN} neurons are activated by noxious stimuli applied to all body parts examined, consistent with studies of electrical activity from unidentified PBN neurons^{12,13}.

To determine whether CGRP^{PBN} neurons code for the intensity of noxious stimuli, we placed the tails of unconscious mice into a temperature-controlled water bath. Neurons were unresponsive to temperatures below 44 °C, but most neurons were activated at 46 °C and above (Fig. 1i). The calcium activity of CGRP^{PBN} neurons increased with increasing temperatures (Fig. 1j and Extended Data Fig. 2a–e), as found previously for unidentified PBN neurons¹². Likewise, CGRP^{PBN} neurons showed an increase in calcium activity as the intensity of electrical shocks delivered to the tail increased (Extended Data Fig. 2f). In awake mice, neurons also exhibited an increase in activity in response to increasing temperatures during a hotplate test (Extended Data Fig. 2g, h), although the mice were able to mitigate some of the pain by alternating the paws that were in contact with the hotplate. We conclude that CGRP^{PBN} neurons encode stimulus intensity.

In awake and anaesthetized mice, electrical shock induced a rapid increase in the calcium activity of most CGRP^{PBN} neurons (Fig. 1k, l and Extended Data Fig. 1d, e). We also measured calcium activity in a neighbouring neuronal population that controls fluid intake—oxytocin receptor-expressing neurons in the dorsal lateral PBN¹⁴.

¹Howard Hughes Medical Institute and Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

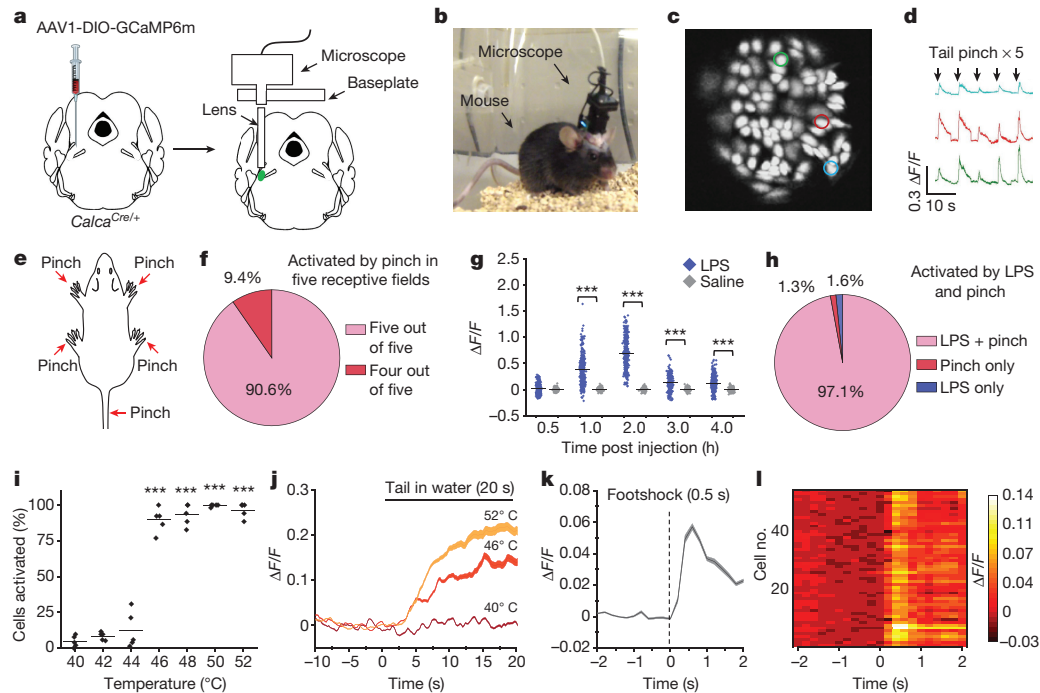


Figure 1 | CGRP^{PBN} neurons are activated by painful stimuli.

a, Schematic⁴⁰ showing configuration for imaging of PBN CGRP neurons. **b**, Mouse with the microscope attached. **c**, Individual neurons identified with software. Coloured circles correspond to traces in **d**. **d**, Calcium fluorescence activity of neurons outlined in **c** recorded during tail pinch. **e**, **f**, The same neurons are activated by pinching the tail or any four paws of anaesthetized mice ($n = 5$ mice, 384 neurons). **g**, Calcium activity following intraperitoneal injection of LPS or saline in awake mice ($n = 6$ mice, 324 neurons saline, 372 neurons LPS); data points are individual neurons and lines are means (two-way ANOVA, Bonferroni's post-hoc). Baseline was 5-min period before injection. **h**, Percentage

of cells activated by LPS and tail pinch. **i**, Percentage of cells activated following placement of tail in heated water bath (40–52 °C) for 20 s ($n = 5$ mice, 286 neurons; one-way repeated measures ANOVA, Tukey's post-hoc). **j**, Time course of calcium activity for temperature gradient study; baseline was 10-s period before water exposure. **k**, Calcium activity time course in response to footshock (0.5 mA, 0.5 s; dashed line) in awake mice ($n = 5$ mice, 317 neurons); baseline was 2-s period before shock. **l**, Plot shows activity of neurons (rows) from a representative mouse that received a footshock. Line graphs are mean \pm s.e.m. (shaded region). $\Delta F/F = (F - F_{\text{baseline}})/F_{\text{baseline}}$. *** $P < 0.001$. For statistical analysis, see Supplementary Information.

Electrical shock did not rapidly affect the activity of those neurons (Extended Data Fig. 1e). The time-locked response of CGRP^{PBN} neurons, but not of neighbouring PBN neurons, further suggests that CGRP^{PBN} neurons are a primary relay for pain signals⁷.

Response to non-painful noxious stimuli

Pain can signal ongoing harm, whereas itch is an adaptive response that may lead to avoidance of harmful plants and insects¹⁵. Although itch-responsive sensory neurons overlap with those that respond to painful stimuli¹⁵, little is known about brain circuits that contribute to itch-related behaviours. We monitored calcium activity in CGRP^{PBN} neurons in anaesthetized mice that received a subcutaneous injection of chloroquine, a drug that causes itching¹⁶. Chloroquine injection increased calcium activity in most CGRP^{PBN} neurons (Fig. 2a).

To determine whether activation of CGRP^{PBN} neurons is required for chloroquine-induced scratching, we injected an AAV expressing Cre-dependent tetanus toxin light chain (AAV-DIO-GFP:TetTox) bilaterally into the PBN of *Calca*^{Cre/+} mice (Fig. 2b). TetTox cleaves synaptobrevin, a synaptic vesicle protein that is necessary for neurotransmitter release¹⁷, thereby preventing neurotransmitter release from Cre-expressing neurons^{7,18}. Compared to mice injected with a GFP control virus, chloroquine-induced scratching was attenuated in mice in which CGRP^{PBN} neurons were silenced (Fig. 2c, d). We examined whether activation of CGRP^{PBN} neurons also contributes to scratching induced by a non-chemical stimulus by placing adhesive stickers on the necks of mice (Fig. 2e). This manipulation activates mechanoreceptors and elicits scratching and paw swipes to remove the sticker, but is not thought to be painful¹⁹. TetTox silencing of CGRP^{PBN} neurons attenuated sticker-removal attempts (Fig. 2f). Therefore, CGRP^{PBN} neurons can be activated by non-painful noxious stimuli and their activation

is required to elicit appropriate behavioural responses. These findings identify a specific population of PBN neurons that contributes to itch-related behaviour²⁰.

CGRP neuron activity during feeding

Because CGRP^{PBN} neurons contribute to satiation⁴, we investigated whether their activity was modulated by feeding-related stimuli in mice that had been deprived of food overnight and then given access to their familiar chow pellets. Calcium activity in CGRP^{PBN} neurons was rapidly inhibited when food was presented (Supplementary Video 2), even before the mice had taken the first bite (Fig. 3a, b and Extended Data Fig. 3a). We also observed rapid decreases in calcium activity before bites (Fig. 3c and Extended Data Fig. 3b). The inhibitory responses were observed in nearly all neurons (Extended Data Fig. 3c, d). To ensure that these effects were not due to movement artefacts, we analysed fluorescence in regions of interest lacking neurons; here we observed no changes before or during feeding (data not shown). We also analysed calcium recordings using constrained non-negative matrix factorization (CNMF)²¹. Individual traces extracted from CNMF analysis show that calcium transients are not synchronized, but all neurons are similarly inhibited upon presentation of food and before bites (Fig. 3d, e and Extended Data Fig. 3e, f). As activation of CGRP^{PBN} neurons suppresses appetite³, their inhibition should support feeding.

We also examined calcium activity as feeding progressed. Food deprivation lowered the activity of CGRP^{PBN} neurons relative to the sated state (Fig. 3f), consistent with increased inhibitory input from AgRP-expressing neurons, which are activated during food deprivation²², and/or decreased excitatory input from anorexigenic neurons in the NTS^{4,5}. After 30 min of feeding, neurons became more active (Fig. 3f), which corresponded to a decrease in food intake

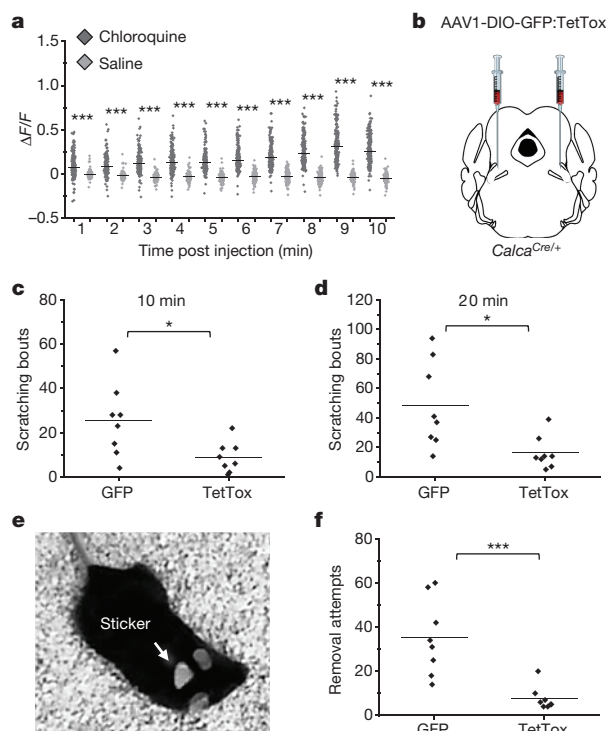


Figure 2 | CGRP^{PBN} neurons are activated by non-painful noxious stimuli. **a**, Calcium fluorescence activity in anaesthetized mice that received subcutaneous injection of chloroquine (8 mM, 10 μ l) or saline control; baseline was 2-min period before injection. Data points are individual neurons and horizontal lines are means ($n = 4$ mice, 231 neurons chloroquine, 216 neurons saline; two-way ANOVA, Bonferroni's post-hoc). **b**, Schematic⁴⁰ showing bilateral PBN injection of *Calca*^{Cre/+} mice with AAV-DIO-TetTox:GFP to prevent neurotransmitter release from CGRP neurons or AAV-DIO-GFP control virus. **c**, **d**, Quantification of chloroquine-induced scratching bouts in GFP and TetTox mice 10 ($P = 0.021$) and 20 min ($P = 0.011$) after injection. **e**, Placement of sticker on neck area of mouse. **f**, Quantification of sticker removal attempts during the 5 min after placing the sticker on GFP and TetTox mice. Graphs (**c**, **d**, **f**) show data from individual mice ($n = 8$ per group; Student's *t*-test, two-tailed); lines represent the mean. * $P < 0.05$, *** $P < 0.001$. For statistical analysis, see Supplementary Information.

compared to the first 30 min (0–30 min, 0.71 ± 0.04 g; 30–60 min, 0.22 ± 0.02 g (mean \pm s.e.m.)). In general, these neurons became more active as feeding progressed (Fig. 3g), compared to either non-fasted or fasted baseline conditions (Fig. 3h). All neurons that were analysed in the feeding study also responded to tail pinch (data not shown).

Coding of food neophobia

Given that CGRP^{PBN} neurons respond to anticipatory signals before food consumption, we investigated whether they respond to other complex sensory inputs by examining the neophobia response to a novel food, an adaptive trait that limits the consumption of unfamiliar, potentially toxic foods. In contrast to their response to the familiar diet, CGRP^{PBN} neurons showed increased calcium activity when mice were presented with an unfamiliar, high-fat food pellet (Extended Data Fig. 4a–f and Supplementary Video 3). CGRP^{PBN} neurons were also activated following each bite during the first trial (Fig. 4a and Extended Data Fig. 4g). However, after mice had become acclimated to the unfamiliar food, CGRP^{PBN} neurons became inhibited before and during consumption of the palatable diet (Fig. 4b, c and Extended Data Fig. 4h, i), as during consumption of their normal diet. The change in the pattern of calcium activity corresponded with a decreased latency to begin consumption during the last trial (20.4 ± 4.1 s; mean \pm s.e.m.) compared to the first trial (115.2 ± 16.0 s). We also monitored calcium activity in response to an inedible novel object (a marble).

CGRP^{PBN} neurons were activated upon first presentation of the object (Supplementary Video 4), a response that dissipated as mice became acclimated to the object (Extended Data Fig. 5). After acclimation, neuronal activity did not decline as with the palatable diet, probably because food is appetitive whereas the marble becomes neutral after acclimation.

We examined the necessity of CGRP^{PBN} neurons for behavioural expression of food neophobia by silencing them with TetTox (Fig. 4d). Compared to control mice, TetTox-injected mice exhibited an attenuated food neophobia response, which was apparent during the first two exposures to a novel palatable diet (Fig. 4e), indicating that CGRP^{PBN} neurons contribute to food neophobia. This result is consistent with attenuation of food neophobia by lesions of the lateral PBN²³.

Reactivation of CGRP^{PBN} neurons by fear memory

With food neophobia, fear behaviour dissipates after repeated exposure, suggesting that a positive memory is formed and that CGRP^{PBN} neurons are sensitive to learning processes. We postulated that CGRP^{PBN} neurons might also respond to a negative memory, such that exposure to a stimulus paired with a painful event would lead to activation of CGRP^{PBN} neurons by that conditioned stimulus.

Mice were subjected to footshock–tone conditioning, in which mice learn to associate a neutral stimulus (tone) with an electrical footshock (Fig. 5a). These mice formed a fear memory, made evident by the fact that the tone alone was sufficient to elicit freezing (a rodent fear response) the next day (Fig. 5b). After conditioning, CGRP^{PBN} neurons were activated during the 10-s tone presentation (Fig. 5c and Extended Data Fig. 6a). Neuronal activation became less robust during extinction with repeated presentations of the tone in the absence of footshock (Fig. 5d, e and Extended Data Fig. 6a), which corresponded with a decrease in freezing behaviour (Fig. 5b). Mice that did not receive footshocks with the tone during conditioning did not exhibit tone-induced freezing or increased calcium activity in CGRP^{PBN} neurons (Fig. 5f–j and Extended Data Fig. 6b). We also investigated whether non-specific contextual cues could activate CGRP^{PBN} neurons in mice that underwent footshock fear conditioning (Extended Data Fig. 6c). Compared to test trials before conditioning, placing mice in the shock chamber after conditioning was sufficient to increase calcium activity in CGRP^{PBN} neurons (Extended Data Fig. 6d–g). Therefore, in addition to being a primary relay of pain signals that are necessary and sufficient to form a fear memory^{7,24}, CGRP^{PBN} neurons also become activated during recall of that aversive memory.

To investigate whether activation of CGRP^{PBN} neurons contributes to learned behavioural fear responses, we subjected mice to footshock fear conditioning (Fig. 6a) followed by CGRP^{PBN} neuron silencing with TetTox after the fear memory had been formed (Fig. 6b). We used greater shock intensities than in the previous experiment to form a more robust fear memory because extinction training occurred weeks after conditioning, to allow sufficient time for TetTox expression. Compared to GFP-injected control mice, mice in which CGRP^{PBN} neurons were silenced rapidly extinguished their freezing response during repeated presentations of the tone without footshocks (Fig. 6c). By the second fear-extinction-training session, fear-conditioned TetTox mice exhibited minimal freezing behaviour (Fig. 6d), similar to control mice that did not undergo fear conditioning (Fig. 6e–h). Conversely, fear-conditioned GFP-control mice continued to freeze in response to the tone during the second extinction-training session (Fig. 6d). Thus, activation of CGRP^{PBN} neurons contributes substantially to fear responses during recall of an aversive memory.

Discussion

Here we show that CGRP^{PBN} neurons become activated in response to danger signals of diverse sensory modalities—visceral and cutaneous pain (chemical, mechanical, thermal and electrical), itch, and meal-related satiety signals—that involve various ascending sensory pathways (trigeminal, spinal and vagal). CGRP^{PBN} neurons are also

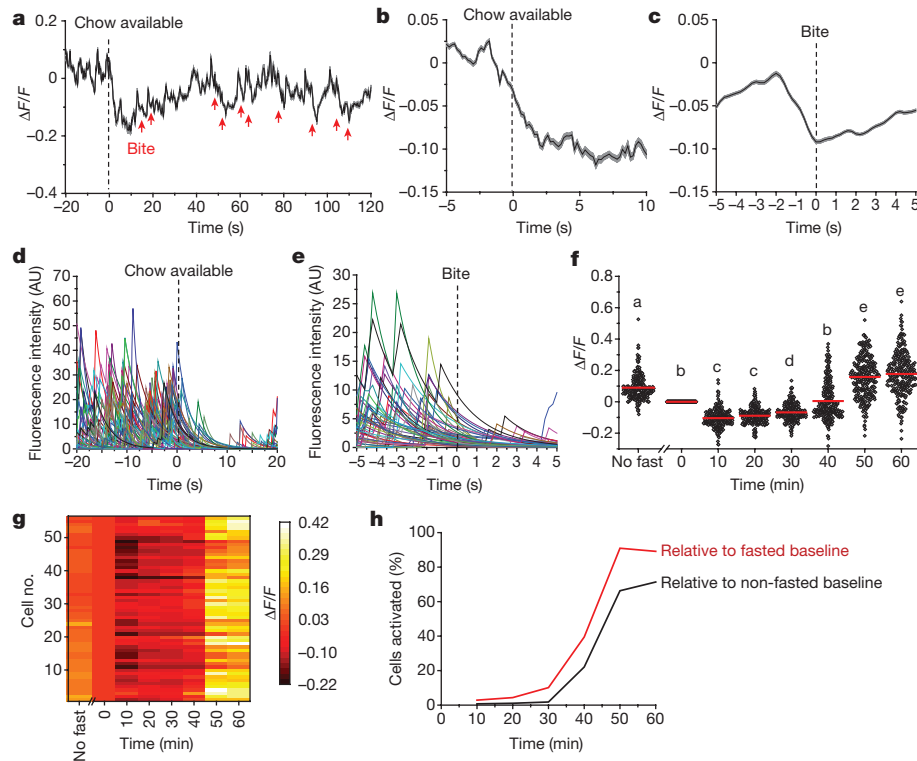


Figure 3 | Rapid control of CGRP^{PBN} neurons before and during feeding. **a**, Average calcium fluorescence activity from a representative mouse (63 neurons) relative to chow being available (dashed line) and each bite taken (red arrows) during a fast-refeed study. **b**, Calcium activity relative to placement of chow in cage ($n = 5$ mice, 282 neurons). Data are mean \pm s.e.m. (shaded region). **c**, Calcium activity relative to taking a bite. Multiple bites were analysed and averaged for each neuron ($n = 5$ mice, 282 neurons). Data are mean \pm s.e.m. (shaded region). **d**, **e**, Forty-seven neuron traces from a representative mouse extracted using CNMF analysis, relative to chow availability (**d**) and to a bite (**e**). **f**, Calcium

activity after refeeding. Data are normalized to time-point 0 min, which was a 60-s baseline before food availability; 'no fast' data were recorded 24 h before fast. Data points are individual neurons and the lines are means ($n = 5$ mice, 282 neurons; one-way repeated measures ANOVA, Tukey's post-hoc). Dissimilar letters above columns of data points indicate statistical differences between time points. **g**, Calcium activity of neurons from a representative mouse after refeeding. **h**, Percentage of neurons activated relative to fasted or non-fasted baseline conditions. For statistical analysis, see Supplementary Information.

activated by danger signals that require cognition, as shown by the food-neophobia and fear-conditioning assays, and TetTox loss-of-function studies reveal that these neurons contribute to both innate

and learned fear behaviours. Thus, CGRP^{PBN} neurons convey negative valence, while parallel pathways are likely to provide information regarding sensory modality and stimulus origin²⁵ to help elicit

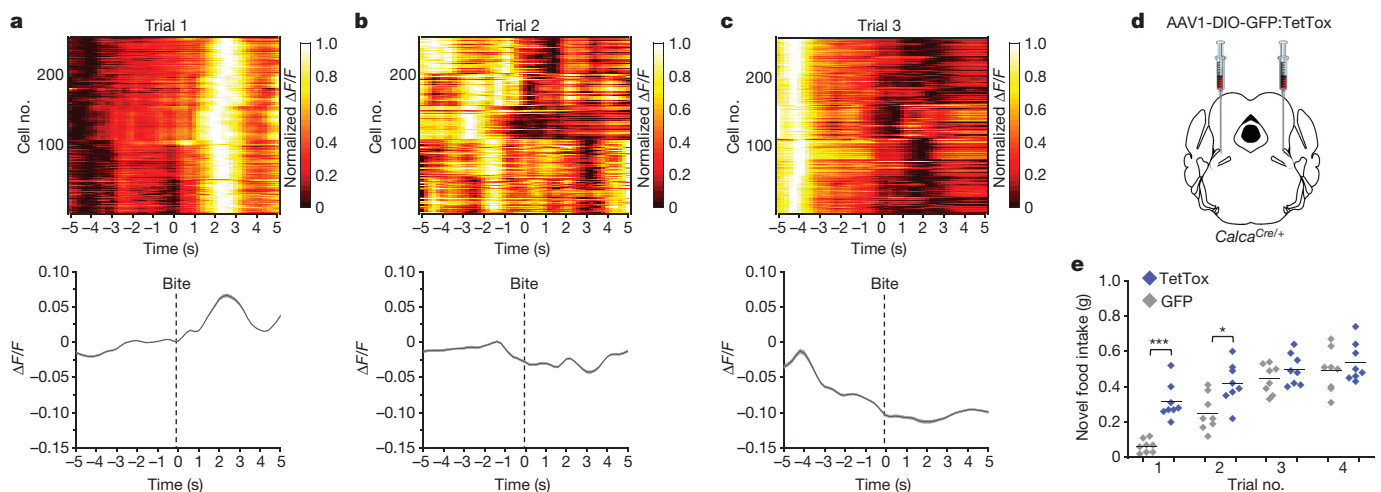


Figure 4 | CGRP^{PBN} neurons are activated in response to a novel and palatable high-fat diet. **a–c**, Calcium fluorescence activity in response to taking a bite from a novel, high-fat diet (HFD) pellet. Line graphs are means \pm s.e.m. (shaded region) for raw $\Delta F/F$ values and plots show normalized $\Delta F/F$ values for neurons (rows) of all experimental mice (trial 1, $n = 5$ mice, 247 neurons; trial 2, $n = 5$ mice, 244 neurons; trial 3, $n = 5$ mice, 254 neurons). Test trials were conducted on consecutive days. **d**, Schematic⁴⁰ showing bilateral PBN injection of *Calca*^{Cre/+} mice with

AAV-DIO-TetTox:GFP (to silence CGRP neurons) or AAV-DIO-GFP control virus. **e**, Non-food-deprived mice received 20 min access to a novel HFD pellet on four consecutive days; data points are individual mice (eight mice per group; two-way repeated measures ANOVA, Bonferroni's post-hoc). Baseline was a 60-s period before HFD pellet was presented. *** $P < 0.001$, * $P < 0.05$. For statistical analysis, see Supplementary Information.

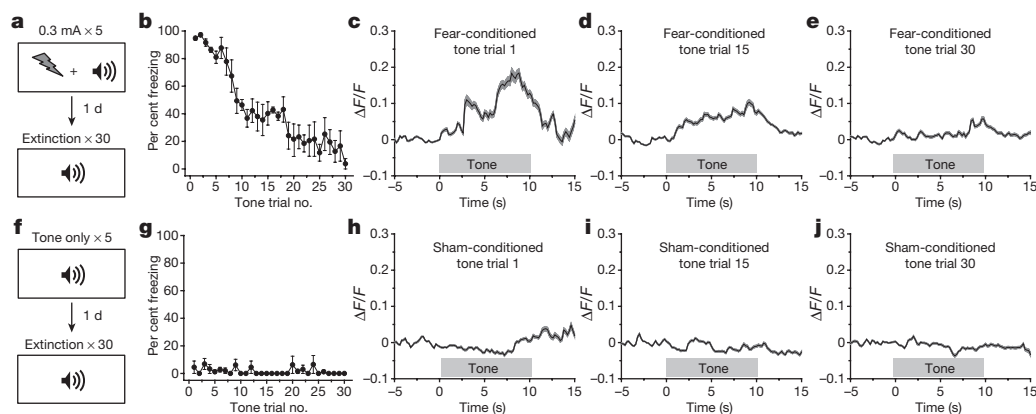


Figure 5 | CGRP^{PBN} neurons are activated during recall of a pain memory. **a**, Fear-conditioning paradigm in which mice received a footshock (0.3 mA; 0.5 s duration) that co-terminated with a 10-s tone (five pairings). **b**, Percentage of time spent freezing (fear-conditioned mice) in response to 10-s tone presentations during the extinction training sessions. **c–e**, Calcium fluorescence activity of neurons in fear-conditioned mice during presentation of tone trials 1, 15, and 30. **f**, Sham-conditioning paradigm. **g**, Percentage of time spent freezing (sham-conditioned mice)

in response to 10-s tone presentations during the extinction training sessions. **h–j**, Calcium activity of neurons in sham-conditioned mice during presentation of tone trials 1, 15, and 30. Fear-conditioned mice, $n = 4$ mice, 214 neurons; sham-conditioned mice, $n = 4$ mice, 169 neurons. Line graphs are mean \pm s.e.m. (shaded region). Baseline was 5-s period before tone presentation. For statistical analysis, see Supplementary Information.

appropriate responses. In addition to the modalities examined here, we hypothesize that CGRP^{PBN} neurons would respond to gustatory²⁶, olfactory²⁷, auditory²⁸, vestibular²⁹, and other visceral stimuli³⁰.

Calcium recordings show that CGRP^{PBN} neurons as a population are activated or inhibited in response to (putative) salient stimuli. This arrangement is also observed in AgRP 'hunger' neurons²²; in that case, separate projections to distinct brain regions contribute to various aspects of hunger-related behaviour and physiology^{31–33}. CGRP^{PBN} neurons project to the central amygdala and BNST^{3,7}, nuclei that mediate behavioural and physiological responses to ingestive and fear-related stimuli^{7,8,34}. Downstream central amygdala neurons express the CGRP receptor⁷, and a subpopulation of those neurons co-express protein kinase C- δ ³⁵, but the identity of downstream BNST neurons is unknown. We postulate that the central amygdala and BNST contribute to different aspects of behaviour and/or alter behaviours over different timescales. Behaviours may also be tuned by the magnitude and duration of CGRP^{PBN} neuronal activation, which could lead to differential activation of postsynaptic targets.

Although we focused here on aversive stimuli, pain-responsive CGRP^{PBN} neurons also respond to feeding-related stimuli. CGRP^{PBN}

neurons are rapidly inhibited before food consumption, indicating that there is bidirectional control of their activity and revealing that they also participate in a normal, non-aversive behaviour. Inhibition from AgRP neurons or decreased input from anorexigenic neurons in the NTS⁵ is likely to dampen CGRP^{PBN} neuron activity during fasting⁴, but the decreases in calcium activity during food consumption are unlikely to come from AgRP neurons because they are inactive during feeding^{22,36}. Because CGRP^{PBN} neurons transmit potential danger signals, their inhibition during starvation might contribute to adaptive behaviours, such as increased risk taking to acquire food³¹, in addition to permitting increased food intake.

Numerous forebrain projections to the PBN^{37,38} are candidates for the control of CGRP^{PBN} neurons in response to sensory signals that require complex processing during neophobia and learning. Notably, fear-conditioned mice with either intact or inhibited CGRP^{PBN} neurons showed similar behavioural responses during the first tone trial of extinction training, but the TetTox-treated mice showed faster extinction of their fear memory. This result suggests that activation of other brain areas is initially sufficient to elicit a learned fear response, but that CGRP^{PBN} neuronal activation is critical for maintaining a persistent

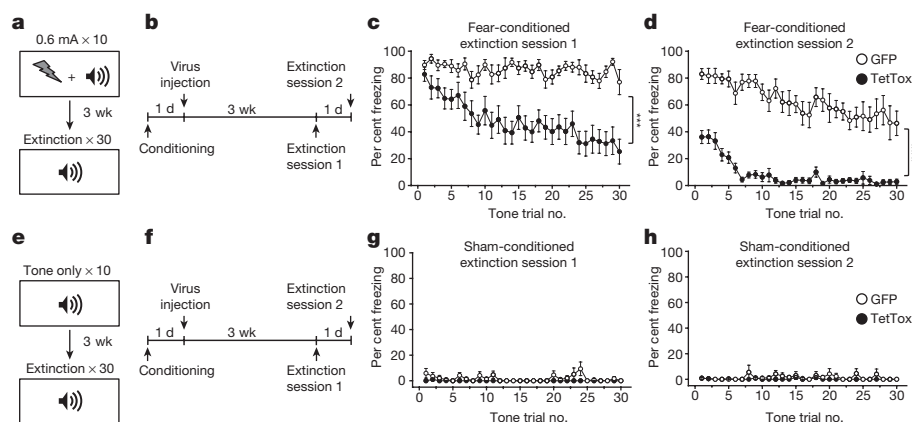


Figure 6 | Silencing of CGRP^{PBN} neurons attenuates fear response during recall of pain memory. **a**, Fear-conditioning paradigm in which mice received a footshock (0.6 mA; 0.5 s duration) that co-terminated with a 10-s tone (ten pairings). **b**, Timeline for experiments and surgeries of mice that underwent fear conditioning. After fear conditioning, mice received bilateral PBN injections of AAV-DIO-TetTox:GFP to inhibit CGRP neurons or AAV-DIO-GFP control virus. **c, d**, Percentage of time spent freezing (fear-conditioned mice) in response to 10-s tone

presentations during extinction sessions ($n = 12$ per group; two-way repeated measures ANOVA). **e**, Sham-conditioning paradigm. **f**, PBN virus injections after sham conditioning. **g, h**, Percentage of time spent freezing (sham-conditioned mice) in response to 10-s tone presentations during extinction training sessions ($n = 6$ per group). Line graphs are mean \pm s.e.m. *** $P < 0.001$. For statistical analysis, see Supplementary Information.

fear state and retarding extinction. Current fear-extinction hypotheses suggest that extinction of a fear memory involves weakening of the original associative memory and/or consolidation of a newly acquired 'safe' memory that overrides the original fear memory³⁹. Our findings are congruent with either interpretation, as activation of CGRP^{PBN} neurons could strengthen the original fear memory or preclude formation of a safe memory.

We expect that future studies will reveal how CGRP^{PBN} neurons are rapidly controlled by various neuronal inputs and shed light on how core brainstem circuits are recruited to produce complex behaviours. These endeavours may ultimately lead to a better understanding of how CGRP^{PBN} neurons alter affective-motivational behaviours associated with transient⁷ and chronic threats⁴⁰, and mental disorders that might involve dysregulation of upstream neurocircuitry.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 March 2017; accepted 9 January 2018.

Published online 21 March 2018.

- Saper, C. B. The house alarm. *Cell Metab.* **23**, 754–755 (2016).
- Saper, C. B. & Lowey, A. D. Efferent connections of the parabrachial nucleus in the rat. *Brain Res.* **197**, 291–317 (1980).
- Carter, M. E., Soden, M. E., Zweifel, L. S. & Palmiter, R. D. Genetic identification of a neural circuit that suppresses appetite. *Nature* **503**, 111–114 (2013).
- Campos, C. A., Bowen, A. J., Schwartz, M. W. & Palmiter, R. D. Parabrachial CGRP neurons control meal termination. *Cell Metab.* **23**, 811–820 (2016).
- Roman, C. W., Derkach, V. A. & Palmiter, R. D. Genetically and functionally defined NTS to PBN brain circuits mediating anorexia. *Nat. Commun.* **7**, 11905 (2016).
- Ritter, R. C. Gastrointestinal mechanisms of satiation for food. *Physiol. Behav.* **81**, 249–273 (2004).
- Han, S., Soleiman, M. T., Soden, M. E., Zweifel, L. S. & Palmiter, R. D. Elucidating an affective pain circuit that creates a threat memory. *Cell* **162**, 363–374 (2015).
- Cai, H., Haubensak, W., Anthony, T. E. & Anderson, D. J. Central amygdala PKC- δ^{+} neurons mediate the influence of multiple anorexigenic signals. *Nat. Neurosci.* **17**, 1240–1248 (2014).
- Lebow, M. A. & Chen, A. Overshadowed by the amygdala: the bed nucleus of the stria terminalis emerges as key to psychiatric disorders. *Mol. Psychiatry* **21**, 450–463 (2016).
- Flusberg, B. A. *et al.* High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nat. Methods* **5**, 935–938 (2008).
- Ghosh, K. K. *et al.* Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).
- Bernard, J. F. & Besson, J. M. The spino(trigemino)pontoamygdaloid pathway: electrophysiological evidence for an involvement in pain processes. *J. Neurophysiol.* **63**, 473–490 (1990).
- Bernard, J. F., Huang, G. F. & Besson, J. M. The parabrachial area: electrophysiological evidence for an involvement in visceral nociceptive processes. *J. Neurophysiol.* **71**, 1646–1660 (1994).
- Ryan, P. J., Ross, S. I., Campos, C. A., Derkach, V. A. & Palmiter, R. D. Oxytocin-receptor-expressing neurons in the parabrachial nucleus regulate fluid intake. *Nat. Neurosci.* **20**, 1722–1733 (2017).
- Braz, J., Solorzano, C., Wang, X. & Basbaum, A. I. Transmitting pain and itch messages: a contemporary view of the spinal cord circuits that generate gate control. *Neuron* **82**, 522–536 (2014).
- Liu, Q. *et al.* Sensory neuron-specific GPCR Mrgprs are itch receptors mediating chloroquine-induced pruritus. *Cell* **139**, 1353–1365 (2009).
- Schiavo, G., Matteoli, M. & Montecucco, C. Neurotoxins affecting neuroexocytosis. *Physiol. Rev.* **80**, 717–766 (2000).
- Kim, J. C. *et al.* Linking genetically defined neurons to behavior through a broadly applicable silencing allele. *Neuron* **63**, 305–315 (2009).
- Ranade, S. S. *et al.* Piezo2 is the major transducer of mechanical forces for touch sensation in mice. *Nature* **516**, 121–125 (2014).
- Mu, D. *et al.* A central neural circuit for itch sensation. *Science* **357**, 695–699 (2017).
- Zhou, P. *et al.* Efficient and accurate extraction of in vivo calcium signals from microendoscope video data. *eLife* (in the press).
- Betley, J. N. *et al.* Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature* **521**, 180–185 (2015).
- Reilly, S. & Trifunovic, R. Lateral parabrachial nucleus lesions in the rat: neophobia and conditioned taste aversion. *Brain Res. Bull.* **55**, 359–366 (2001).
- Carter, M. E., Han, S. & Palmiter, R. D. Parabrachial calcitonin gene-related peptide neurons mediate conditioned taste aversion. *J. Neurosci.* **35**, 4582–4586 (2015).
- Tovote, P., Fadok, J. P. & Lüthi, A. Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
- Ricardo, J. A. & Koh, E. T. Anatomical evidence of direct projections from the nucleus of the solitary tract to the hypothalamus, amygdala, and other forebrain structures in the rat. *Brain Res.* **153**, 1–26 (1978).
- Day, H. E. W., Masini, C. V. & Campeau, S. The pattern of brain c-fos mRNA induced by a component of fox odor, 2,5-dihydro-2,4,5-trimethylthiazoline (TMT), in rats, suggests both systemic and processive stress characteristics. *Brain Res.* **1025**, 139–151 (2004).
- Illing, R. B., Michler, S. A., Kraus, K. S. & Laszig, R. Transcription factor modulation and expression in the rat auditory brainstem following electrical intracochlear stimulation. *Exp. Neurol.* **175**, 226–244 (2002).
- Suzuki, T., Sugiyama, Y. & Yates, B. J. Integrative responses of neurons in parabrachial nuclei to a nauseogenic gastrointestinal stimulus and vestibular stimulation in vertical planes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **302**, R965–R975 (2012).
- Kaur, S. *et al.* Glutamatergic signaling from the parabrachial nucleus plays a critical role in hypercapnic arousal. *J. Neurosci.* **33**, 7627–7640 (2013).
- Padilla, S. L. *et al.* Agouti-related peptide neural circuits mediate adaptive behaviors in the starved state. *Nat. Neurosci.* **19**, 734–741 (2016).
- Steculorum, S. M. *et al.* AgRP neurons control systemic insulin sensitivity via myostatin expression in brown adipose tissue. *Cell* **165**, 125–138 (2016).
- Betley, J. N., Cao, Z. F., Ritola, K. D. & Sternson, S. M. Parallel, redundant circuit organization for homeostatic control of feeding behavior. *Cell* **155**, 1337–1350 (2013).
- Kim, S. Y. *et al.* Diverging neural pathways assemble a behavioural state from separable features in anxiety. *Nature* **496**, 219–223 (2013).
- Kim, J., Zhang, X., Muralidhar, S., LeBlanc, S. A. & Tonegawa, S. Basolateral to central amygdala neural circuits for appetitive behaviors. *Neuron* **93**, 1464–1479.e5 (2017).
- Chen, Y., Lin, Y. C., Kuo, T. W. & Knight, Z. A. Sensory detection of food rapidly modulates arcuate feeding circuits. *Cell* **160**, 829–841 (2015).
- Tokita, K., Inoue, T. & Boughter, J. D. Jr. Afferent connections of the parabrachial nucleus in C57BL/6J mice. *Neuroscience* **161**, 475–488 (2009).
- Zséfi, G. *et al.* Elucidation of the anatomy of a satiety network: Focus on connectivity of the parabrachial nucleus in the adult rat. *J. Comp. Neurol.* **524**, 2803–2827 (2016).
- Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* **12**, 120–150 (2007).
- Campos, C. A. *et al.* Cancer-induced anorexia and malaise are mediated by CGRP neurons in the parabrachial nucleus. *Nat. Neurosci.* **20**, 934–942 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements Research was supported by a fellowship grant from Hope Funds for Cancer Research (C.A.C.), a National Institutes of Health (NIH) training grant (C.W.R., T32DK007247), and an NIH grant (R.D.P., R01-DA24908). Inscopix provided the calcium imaging equipment via their DECODE program. We thank C. de Solages and L. Cardy (Inscopix) for advice regarding calcium imaging, Y. S. Jo for help with conditioning equipment, M. Chiang for maintaining the mouse colony, and S. Han for making the GCaMP6m virus.

Author Contributions C.A.C. designed and conducted the calcium imaging studies. A.J.B. performed the stereotaxic surgeries for loss-of-function studies. C.A.C. and A.J.B. designed and conducted the fear studies. C.A.C. and C.W.R. designed and conducted the itch studies. R.D.P. provided guidance and resources. C.A.C. wrote the manuscript with input from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.A.C. (camposca@uw.edu) and R.D.P. (palmiter@uw.edu).

Reviewer Information Nature thanks C. Alexander and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

All experimental protocols were conducted according to US National Institutes of Health guidelines for animal research and approved by the Institutional Animal Care and Use Committee at the University of Washington.

Animals. Heterozygous *Calca*^{Crel/+} mice and *Oxtr*^{Crel/+} mice (C57Bl/6 background) were generated and maintained as previously described^{3,14}. Male mice, 3–6 months of age, were used for all studies. Mice from each litter were randomly assigned into treatment groups. Following stereotaxic surgery, mice were singly housed for at least three weeks before and during experimentation, with *ad libitum* access (unless noted otherwise) to standard chow diet (LabDiet 5053) in temperature- and humidity-controlled facilities with 12-h light–dark cycles. Behavioural experiments were conducted during the light cycle, starting at approximately 3 h into the light cycle. No statistical methods were used to predetermine sample size. Unless stated otherwise, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Virus production. AAV1-CBA-DIO-GCaMP6m, AAV1-CBA-DIO-GFP:TetTox, and AAV1-EF1a-DIO-GFP viral vectors were produced by transfecting HEK cells and then purifying cell extracts by pelleting through sucrose and by CsCl gradient ultracentrifugation. Final pellets were suspended in 0.1 M PBS. Viruses were titred as described⁴. In pilot experiments, we injected varying concentrations of AAV1-CBA-DIO-GCaMP6m, examined histological sections four weeks after virus injection, and then chose the highest virus titre that produced GCaMP6m expression that was limited to the soma.

Stereotaxic surgeries. All mice were anaesthetized with 1.5 to 2.0% isoflurane (mixed with 0.8 l per min of oxygen) for stereotaxic surgeries and virus injections. For calcium imaging, *Calca*^{Crel/+} mice received a unilateral PBN injection of AAV1-DIO-GCaMP6m (0.5 µl). Virus was injected using a glass capillary attached to a Nanoject (Drummond, #3-00-204) using the following coordinates: −4.65 mm posterior to bregma, −1.30 mm lateral from midline, and −2.90 mm ventral to dura. These coordinates were verified with Fluoro-Gold (Sigma-Aldrich, #39286) injections in non-experimental mice, which produced Fluoro-Gold labelling that was restricted to the lateral PBN. AAV1-DIO-GFP and AAV1-DIO-GFP:TetTox viruses (for loss-of-function studies) were injected bilaterally.

Mice were prepared for calcium imaging as described⁴¹. In brief, three weeks after AAV1-CBA-DIO-GCaMP6m virus injection, *Calca*^{Crel/+} mice were anaesthetized (as described above) and implanted with a microendoscope lens (6.1 mm length, 0.5 mm diameter; Inscopix, #100-000588) with the assistance of a ProView implant kit (Inscopix, #100-000754) that allowed us to visualize fluorescent activity during the lens implant. The lens was targeted to be ~200–300 µm above the neurons using the following coordinates: −4.80 mm posterior to bregma, −1.70 mm lateral from midline, and −3.20 to −3.40 mm ventral to skull surface. Some lenses were purposely implanted deeper to record from ventral CGRP^{PBN} neurons. To visualize oxytocin receptor neurons, the lens was targeted more dorsal and medial: −4.80 mm posterior to bregma, −1.40 mm lateral from midline, and −3.00 mm ventral to skull surface. One week after lens implantation, mice were anaesthetized and a baseplate (Inscopix, #100-000279) was implanted above the lens. The baseplate provides an interface for attaching the miniature microscope during calcium-imaging experiments, but during other times a baseplate cover (Inscopix, #100-000241) was attached to prevent damage to the microendoscope lens. Post-hoc histological analysis showed variability in the lens placement relative to the rostral–caudal extent of the PBN, suggesting that CGRP^{PBN} neurons in different parts of the PBN were sampled (Extended Data Fig. 7). Out of approximately 60 mice that were injected with GCaMP6m virus, 20 had successful lens implantation and were used for studies.

Calcium imaging studies. All calcium imaging was recorded at 5 frames per second, 200-ms exposure time, and 10–30% LED power using a miniature microscope from Inscopix (nVista). The recording parameters were based on pilot studies that demonstrated minimal photo-bleaching using these settings. In our pilot studies we also recorded calcium activity with higher frame rates (10 and 15 frames per second), but we were unable to find an optimal virus titre that produced sufficient fluorescent activity to record at high frame-rates while maintaining the ability to discriminate individual neurons. CGRP^{PBN} neurons are densely packed in a small brain area and higher GCaMP6m virus titres produced dense dendritic labelling that overlapped with the soma of neighbouring neurons. Ethovision (Noldus, XT 10) was used to trigger and synchronize behavioural video recordings with calcium recordings.

Order of experiments. In one cohort of *Calca*^{Crel/+} mice ($n = 5$), fast and refeeding studies were conducted first, followed by neophobia studies, and then studies involving noxious stimuli (pinch, shock, LPS, itch, and temperature studies). A second cohort of naive *Calca*^{Crel/+} mice were used for fear conditioning studies ($n = 4$ per group). A third group of *Calca*^{Crel/+} mice ($n = 3$) were used to examine graded shock intensities. *Oxtr*^{Crel/+} mice ($n = 3$) were used for a shock study.

A fourth group of *Calca*^{Crel/+} mice ($n = 3$) were used for the hotplate test and contextual fear conditioning.

Fasting–refeeding. Feeding studies (including food neophobia experiments) were conducted before the pain studies. Mice were acclimated to handling and attachment of the miniature microscope for approximately one hour per day for two weeks. After the acclimation period, mice were food deprived overnight, and a standard chow pellet (their normal food) was placed on the floor of their cage following a 1-min baseline recording period. Calcium activity was recorded continuously for 5 min after receiving access to food, followed by 2-min duration recordings 10, 20, 30, 40, 50, and 60 min after refeeding. For analysis of fluorescent activity relative to taking a bite from the food pellet, we analysed bites during the first 2 min of fasting–refeeding, and averaged these values for each neuron. The timing of bites was confirmed visually using the behavioural videos.

Food neophobia. After a baseline recording period (1 min), non-food-deprived mice were given access to a novel, palatable food pellet (45 kcal% high-fat diet; Research Diets, #D12451). Calcium activity was recorded continuously for 10 min after the onset of access to the novel food. Mice had access to the food pellet for 20 min. Bites were confirmed using visual confirmation of the behavioural videos. Test trials were conducted on consecutive days.

Novel object. We used the same experimental design as with the food neophobia studies, except a dark blue marble (1 cm diameter) was placed in the cage instead of a novel food pellet.

Cutaneous pain. For analysis of calcium activity in response to pinch, mice were anaesthetized with 1.5–2.0% isoflurane (mixed with 0.8 l per min oxygen). Once the calcium activity was stable under baseline conditions, mice received five tail pinches using forceps (5-s duration, 2-min interval between pinches). In a separate recording session, we analysed calcium activity in response to pinch of the tail and paws (5-s duration, 2-min interval between pinches). In another recording session, the tail was placed in a temperature-controlled water bath ranging from 40°C to 52°C (20-s duration, 3-min interval between each temperature gradient). Calcium responses to the tail being placed in 56°C and 60°C water (10-s duration, 5-min interval between trials) were examined on a separate day. To examine calcium responses to facial stimuli, a heated metal rod (probing surface area ~1 cm²) at ~52°C (the metal rod was in 52°C water for 10 min) was placed on the lip area for 20 s. Pinches and thermal stimuli were manually synchronized to calcium recordings by applying the stimulus at a designated time during the recording. Mice underwent a 0.5-mA footshock (0.5-s duration) that was synchronized to the calcium imaging recordings by triggering the shock with the calcium imaging acquisition software (Inscopix, nVista HD) and Ethovision (Noldus, XT 10). For tail shock, electrode gel (Spectra 360, Parker Laboratories) was placed on the tail and the positive and negative leads (18-ga copper wire) from an electrical stimulator (model #A320D, World Precision Instruments) were placed 1 cm apart. A transistor-transistor logic (TTL) pulse to the stimulator was used to synchronize calcium recordings with delivery of an electrical current (2–6 mA, 2 s duration). The hotplate test involved removing the mouse from its home cage and placing it onto a pre-heated aluminium plate (15 × 15 cm) enclosed within a transparent Plexiglas chamber that prevents the mouse from escaping (Coulbourn Instruments). We examined temperatures between 40°C and 56°C (20-s duration, 3-min interval between each temperature gradient, animal placed in home cage during 3-min interval). Except for the footshock and hotplate studies, all experiments were conducted under anaesthesia. Recording sessions were conducted once every other day.

Visceral pain. After a 5-min baseline recording period, awake and freely moving mice received an intraperitoneal injection of lipopolysaccharide (LPS; 50 µg/kg; Calbiochem, #437650). Calcium activity was recorded 0.5, 1, 2, 3, and 4 h after injection (2 min duration for each time point). Using a similar experimental design, mice received a control saline (0.9%) injection on a separate day.

Chloroquine-induced itch. Mice were anaesthetized with 1.5–2.0% isoflurane (mixed with 0.8 l per min oxygen). Once the calcium activity was stable under baseline conditions, mice received either a chloroquine (10 µl; 8 mM dissolved in 0.9% saline; Acros Organics) or saline control injection with a Hamilton syringe (26 s ga; #204001). After injection, calcium activity was monitored for 10 s every minute for 10 min.

Footshock fear conditioning. A separate cohort of animals was used for the fear-conditioning studies. After acclimation to handling and to the miniature microscope, mice underwent footshock fear conditioning in shock chamber (Med Associates), which involved five cue pairings (3 kHz, 80 dB, 10-s duration) that co-terminated with a footshock (0.3 mA, 0.5-s duration, randomized inter-trial interval with a range of 70–140 s and average of 90 s). Sham-conditioned control mice did not receive a footshock. One week later, mice were fitted with the miniature microscope for calcium recordings, and underwent fear extinction in a different but similar-sized chamber (white walls and floor instead of shock

chamber that had Plexiglas walls and shock grid floor). During the extinction training session, mice received 30 tone cues (in the absence of footshock) using the same intertrial interval parameters as during conditioning. A TTL pulse to the shock chamber (Med Associates) was used to trigger and synchronize tones with calcium recordings. Time spent freezing (during the tone), defined as immobility up until any movement of the head or body, was manually scored with a stopwatch (experimenter was blind to treatments). For contextual fear conditioning, mice were placed in the shock chamber for 15 min on two consecutive days before conditioning to measure baseline calcium activity. On the third day, mice were placed in the shock chamber and received 10 footshocks (0.6 mA, 0.5-s duration, randomized intertrial interval with a range of 70–140 s and average of 90 s). On two consecutive days after conditioning, mice were placed in the same shock chamber (shock grids exposed) for 15 min each day to measure calcium activity after fear conditioning. Calcium activity was recorded during the first 5 min of each trial.

Calcium image processing. Calcium recording files were down-sampled (spatial binning factor of 2) to reduce processing time and file size, followed by correction of rigid brain movement using the motion correction application (Inscopix, Mosaic v1.2). For analysis of recordings conducted in anaesthetized mice, individual component analysis and principle component analysis (ICA/PCA) applications were used to identify individual cells and to extract their respective calcium traces. Because the microendoscope lens was implanted between the cerebellum and brainstem, this resulted in a slightly unstable lens and non-rigid motion that could not be corrected with recordings from awake mice. Consequently, we were unable to use ICA/PCA to extract traces from individual cells; we manually drew small regions of interest to extract calcium traces from cells. Raw traces were converted to $\Delta F/F$ ($F - F_{\text{baseline average}}/F_{\text{baseline average}}$), where F was the fluorescence at any given time-point and $F_{\text{baseline average}}$ was the average fluorescence during a designated baseline period. Additional analyses were conducted using CNMF for microendoscopic data²¹.

CGRP^{PBN} TetTox studies. *Chloroquine-induced scratching.* Approximately one month after bilateral PBN injections of AAV1-DIO-GFP:TetTox or AAV1-DIO-GFP, mice were habituated to the experimental paradigm for two consecutive days. On the test day, mice were placed in a cage with a small amount of bedding for 40 min, followed by a subcutaneous injection of chloroquine (10 μ L, 8 mM in 0.9% saline) in the nape of the neck while scruffing the animal. After returning the animal to the cage, behaviour was videotaped for 20 min and scratching bouts were analysed. One or many consecutive scratches with any paw were counted as a single bout. Scratches interrupted by the animal licking the paw used for scratching or putting the paw back on the floor of the cage were counted as separate bouts.

Adhesive sticker removal test. The same experimental cohort of animals as was used to examine chloroquine-induced scratching was also used to determine whether inhibition of CGRP^{PBN} neurons prevents sticker-induced scratching. One week after the itch study, mice were placed in a cage with a small amount of bedding for 5 min before a red adhesive sticker (1/4 of a Shamrock brand dot label; #SDL-75-21 dark red) was placed on the back of the animal (near the nape of the neck). After the animal was returned to the cage, behaviour was videotaped for 5 min. This manipulation induced scratching (similar to chloroquine injection) and behind-

the-ear swipes with the front paw(s). A scratching bout or a behind-the-ear swipe was counted as a single removal attempt.

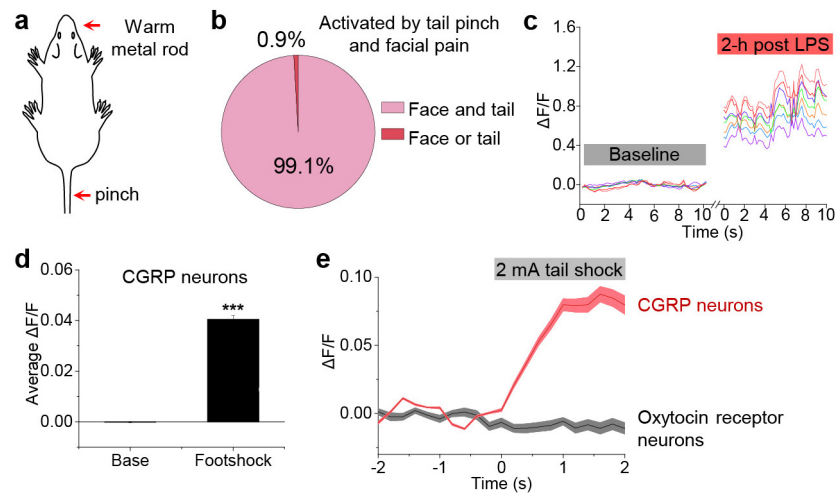
Food neophobia. One group of *Calca^{Grel/+}* mice received bilateral PBN injections of AAV1-DIO-GFP:TetTox to inhibit CGRP^{PBN} neurons, whereas a second group received control injections of AAV1-DIO-GFP. Three weeks after surgery, including a one-week period of acclimation to handling, non-food-deprived mice were presented with a novel and palatable food pellet (45 kcal% high-fat diet) for 20 min. This study was repeated three more times on consecutive days. Mice were not food deprived before the study and they had *ad libitum* access to standard chow and water during the experiment.

Fear conditioning. Prior to fear conditioning or sham conditioning, mice were randomly assigned to the following groups: TetTox fear conditioning, GFP fear conditioning, TetTox sham conditioning, and GFP sham conditioning. Mice in fear conditioning groups underwent footshock fear conditioning, which involved 10 tone pairings (3 kHz, 80 dB, 10-s duration) that co-terminated with a footshock (0.6 mA, 0.5-s duration, randomized intertrial interval with a range of 70–140 s and average of 90 s). Sham-conditioned control mice did not receive a footshock. After the conditioning phase, mice received bilateral PBN injections of AAV1-DIO-GFP:TetTox or AAV1-DIO-GFP (based on their assignment before conditioning). Three weeks later, mice underwent fear extinction training (as described in the calcium imaging study), followed by a second extinction training session the following day. We chose this conditioning paradigm (0.6 mA shock, 10 pairings compared to 0.3 mA, 5 pairings in calcium imaging study) because the extinction training occurred three weeks after conditioning (as opposed to one day for the calcium imaging study) to allow for expression of the virus. Fear extinction training was video recorded and analysed by an investigator who was blind to the treatment. Freezing was defined as complete immobility, including absence of movement of the head and nose.

Statistical analysis. Data were analysed using OriginPro 2016 (OriginLab). For calcium imaging studies, sample sizes were based on published studies^{22,36}. Neurons were considered activated if stimuli produced an increase in fluorescent activity three s.d. above baseline fluorescent activity. For CGRP^{PBN} neuron inhibition studies, post-hoc histological analysis showed bilateral transduction of CGRP^{PBN} neurons with GFP:TetTox or GFP; therefore, all animals were included for final analyses. For graphs comparing multiple treatments (one time-point), we used repeated-measures one-way ANOVA tests, with Tukey's post hoc multiple comparisons test. Data sets with multiple treatments and time-points were analysed with repeated-measures two-way ANOVA tests (time repeated factor), followed by Bonferroni's post hoc tests. Two-tailed Student's *t*-tests were used where appropriate.

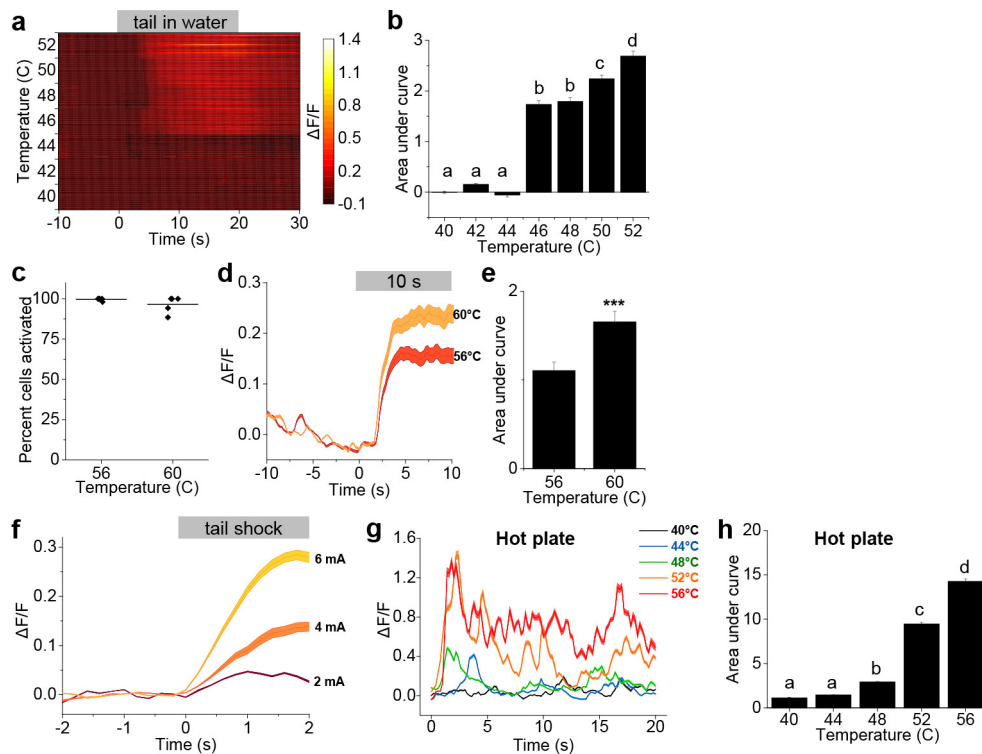
Data availability. The datasets generated during the current study are available from the corresponding author on reasonable request.

- Resendez, S. L. *et al.* Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. *Nat. Protocols* **11**, 566–597 (2016).



Extended Data Figure 1 | CGRP^{PBN} neurons are activated by facial pain and electrical shock. **a, b**, Percentage of cells activated by tail pinch and placing a warm metal rod on the lip of anaesthetized mice ($n = 6$ mice, 397 neurons). **c**, Representative calcium traces of the same individual CGRP neurons during baseline and 2 h after LPS injection in awake mice. **d**, Average calcium fluorescence activity 1 s before (Base) and during a footshock (0.5 mA, 0.5 s duration; $n = 5$ mice, 317 neurons); bar graph is mean \pm s.e.m.. **e**, Calcium activity time course of CGRP neurons ($n = 3$ mice, 158 neurons) and oxytocin receptor neurons ($n = 3$ mice,

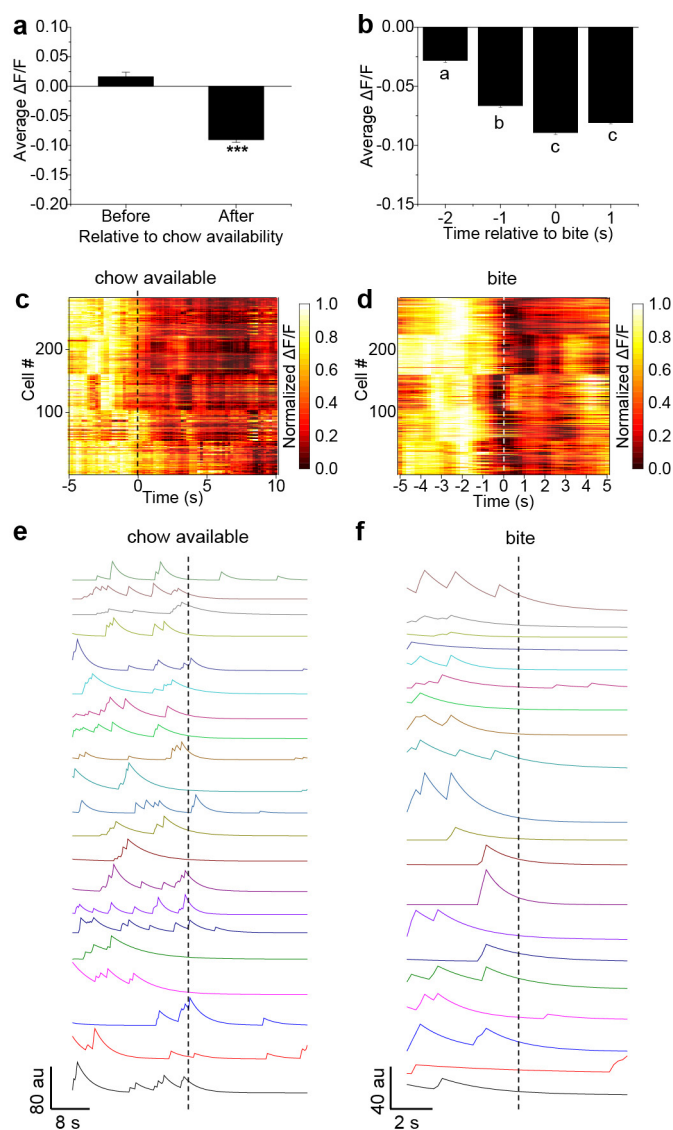
65 neurons) in response to electrical tail shock (2 mA, 2 s duration) in anaesthetized mice; baseline was 2-s period before shock. AAV-DIO-GCaMP6m was injected into the lateral PBN of *Calca*^{Cre/+} and *Oxt*^{Cre/+} mice. For imaging oxytocin receptor neurons, the lens was placed over the dorsal lateral PBN, in contrast to external lateral PBN targeting for CGRP neurons. Line graph is mean \pm s.e.m. (shaded region). *** $P < 0.001$, paired Student's *t*-test (two-tailed). Related to Fig. 1. For statistical analysis, see Supplementary Information.



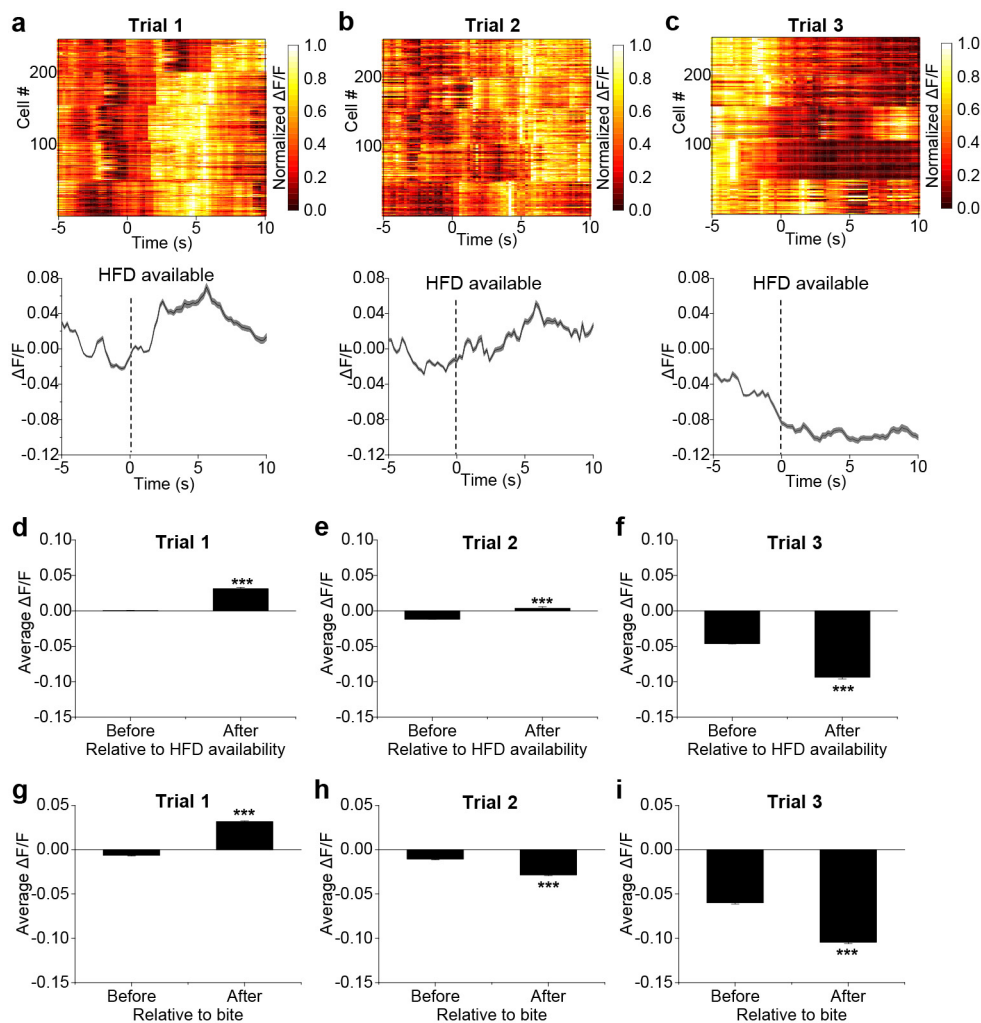
Extended Data Figure 2 | CGRP^{PBN} neurons encode stimulus intensity.

a, Plot shows calcium fluorescence activity of neurons in an anaesthetized mouse following placement of the tail in a water bath ranging from 40°C to 52°C ($n = 56$ neurons). **b**, Area under the curve of calcium activity following placement of tail in 40–52°C water bath ($n = 5$ mice, 286 neurons). **c**, Percentage of cells activated by placement of tail in 56 or 60°C water in anaesthetized mice; data points are individual mice ($n = 4$ mice) and line represents mean. **d**, Time course and peak neuronal responses to tail being placed in 56 or 60°C water ($n = 4$ mice, 111 neurons). Baseline was 10-s period before placing tail in water. **e**, Area under the curve of calcium activity following placement of tail in water bath. **f**, Time course and peak neuronal responses to tail receiving graded

intensities of electrical current (2–6 mA, 2 s duration; $n = 3$ mice, 105 neurons) in anaesthetized mice. Baseline was 2-s period before shock. **g**, Time course and peak calcium activity of neurons in an awake mouse (46 neurons) during exposure to a hotplate (40–56°C, 20-s duration). Baseline was 2-min period in home cage before the hotplate test. **h**, Area under the curve of calcium activity following placement of awake mice on the hotplate ($n = 3$ mice, 126 neurons). Bar graphs are mean \pm s.e.m.; line graphs are mean \pm s.e.m. (shaded region). *** $P < 0.001$; dissimilar letters above columns indicate statistical differences between treatments. **b**, **h**, One-way repeated measures ANOVA, Tukey's post-hoc; **e**, paired Student's t -test (two-tailed). Related to Fig. 1. For statistical analysis, see Supplementary Information.

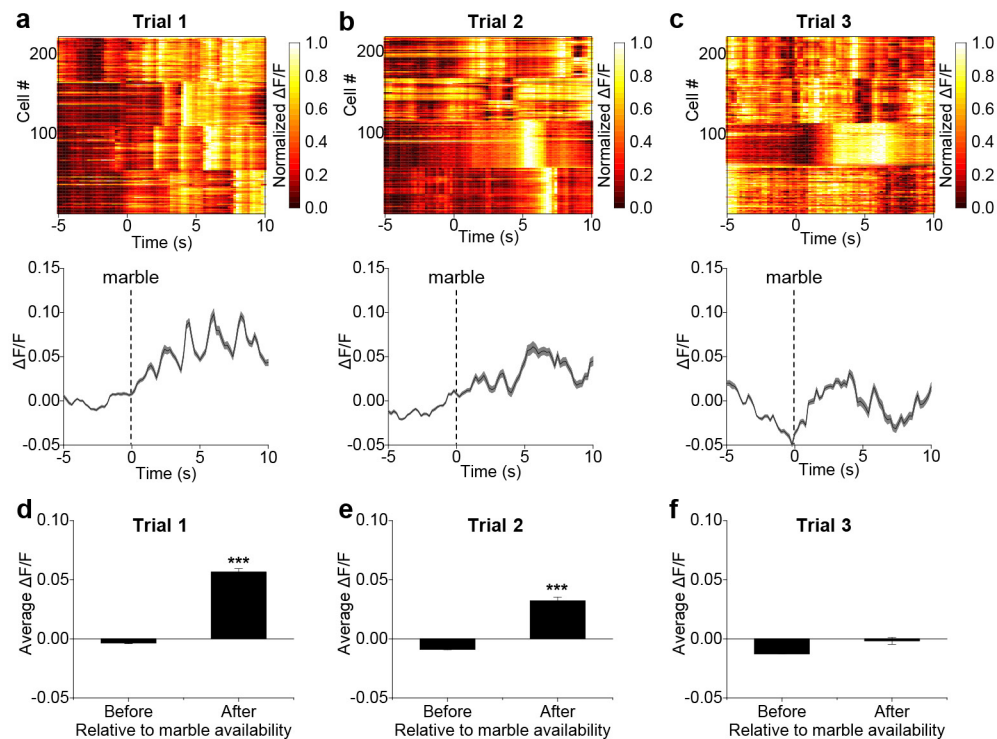


Extended Data Figure 3 | Additional analysis of calcium imaging from fast-refeed study. **a**, Average calcium fluorescence activity 5 s before and 5 s after presentation of standard chow pellet ($n = 5$ mice, 282 neurons); mean \pm s.e.m.. **b**, Calcium activity relative to taking a bite ($n = 5$ mice, 286 neurons). **c**, **d**, Plot showing calcium activity (normalized $\Delta F/F$) relative to placement of chow in cage (**c**) and relative to bite (**d**). Plots show neurons (rows) from all experimental mice ($n = 5$ mice, 282 neurons). **e**, **f**, Representative calcium traces extracted using CNMF analysis, relative to chow availability (**e**) and taking a bite (**f**). Baseline was 60-s period before chow availability for all data shown. *** $P < 0.001$; dissimilar letters above columns indicate statistical differences between time points; **a**, paired Student's t -test (two-tailed); **c**, one-way repeated measures ANOVA, Tukey's post-hoc. Related to Fig. 3. For statistical analysis, see Supplementary Information.



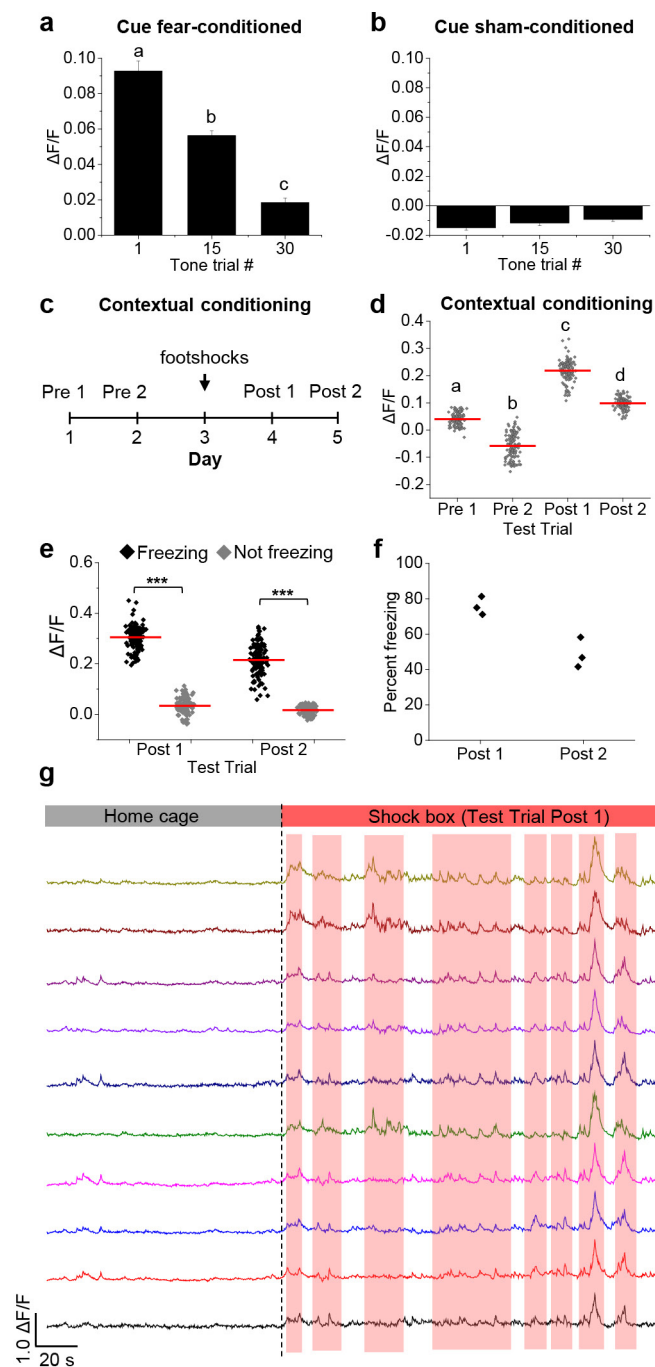
Extended Data Figure 4 | Additional analysis of calcium imaging from food neophobia studies. **a–c**, Change in calcium fluorescence activity in response to a novel, high-fat diet (HFD) pellet. Test trials were conducted on consecutive days. Line graphs are means \pm s.e.m. (shaded region) for raw $\Delta F/F$ values and plots show normalized $\Delta F/F$ values for neurons (rows) of all experimental mice. **d–f**, Average calcium activity 5 s before and 5 s after presentation of HFD pellet during three test trials.

g–i, Average calcium activity 5 s before and 5 s after taking a bite from the HFD pellet. Bar graphs are means \pm s.e.m. Test trial 1, $n = 5$ mice, 247 neurons; trial 2, $n = 5$ mice, 244 neurons; trial 3, $n = 5$ mice, 254 neurons. Baseline was 60-s period before HFD pellet was presented. *** $P < 0.001$; **d–i**, paired Student's t -test (two-tailed). Related to Fig. 4. For statistical analysis, see Supplementary Information.

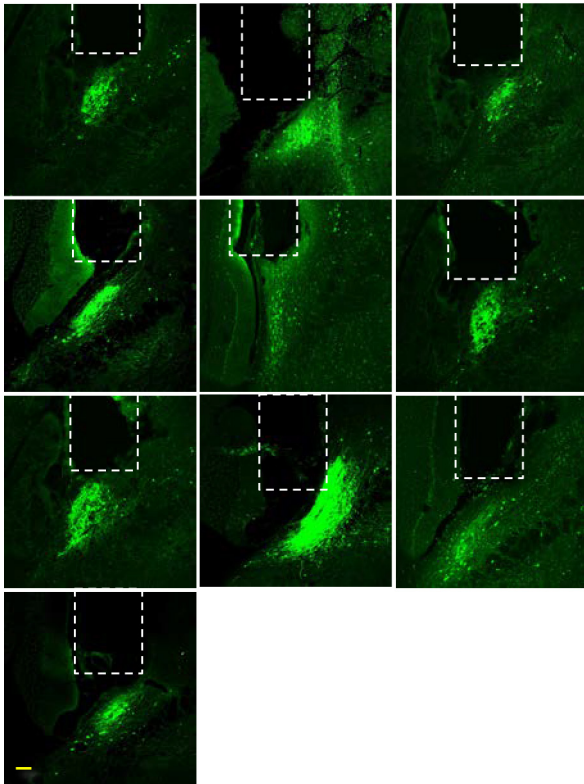


Extended Data Figure 5 | CGRP^{PBN} neurons are activated during exposure to a novel marble. **a–c,** Change in calcium activity in response to a marble being placed in home cage. Line graphs are means \pm s.e.m. (shaded region) for raw $\Delta F/F$ values and plots show normalized $\Delta F/F$ values for neurons (rows) of all experimental mice. **d–f,** Average calcium activity 5 s before and 10 s after placing a marble in the cage. Graphs are

means \pm s.e.m. Test trial 1, $n = 4$ mice, 220 neurons; trial 2, $n = 4$ mice, 218 neurons; trial 3, $n = 4$ mice, 222 neurons. Baseline was 60-s period before marble was presented. *** $P < 0.001$; **d–f,** paired Student's t -test (two-tailed). Related to Fig. 4. See also Supplementary Video 4. For statistical analysis, see Supplementary Information.



Extended Data Figure 6 | Additional calcium imaging analysis from fear recall studies. **a, b**, Average calcium fluorescence activity during 10-s tone presentations in fear-conditioned (**a**) and sham-conditioned (**b**) mice during tone trials 1, 15, and 30. Relative fluorescence was calculated from a 5-s baseline period before tone presentation. Graphs are mean \pm s.e.m.; fear-conditioned mice, $n = 4$ mice, 214 neurons; sham-conditioned mice, $n = 4$ mice, 169 neurons. **c**, Contextual fear conditioning paradigm in which mice were placed in a shock chamber before (Pre 1 and Pre 2) and after (Post 1 and Post 2) footshock conditioning. **d**, Calcium activity of individual neurons (horizontal lines are means) during a 5-min exposure to the shock chamber before and after fear conditioning (Pre 1, $n = 3$ mice, 120 neurons; Pre 2, $n = 3$ mice, 125 neurons; Post 1, $n = 3$ mice, 124 neurons; Post 2, $n = 3$ mice, 121 neurons). Relative fluorescence was normalized to a 2-min baseline recording period. **e**, Calcium activity of individual neurons (horizontal lines are means) during freezing and non-freezing behaviour after conditioning. **f**, Percentage of time spent freezing during exposure to the shock chamber after conditioning. Data points represent individual mice ($n = 3$). **g**, Representative individual neuron traces showing calcium activity while in home cage versus shock chamber after contextual fear conditioning. The dotted line denotes when the shock chamber recording began. The shaded regions annotate when the mouse was freezing. Dissimilar letters above columns indicate statistical differences between trials. *** $P < 0.001$; **a, d, e**, One-way repeated measures ANOVA, Tukey's post-hoc. Related to Fig. 5. For statistical analysis, see Supplementary Information.



Extended Data Figure 7 | Lens placement over CGRP^{PBN} neurons. Representative images of lens placement for imaging CGRP neurons (green). Dashed lines are approximations of lens placement. Scale bar is 100 μm .

Extensive impact of non-antibiotic drugs on human gut bacteria

Lisa Maier^{1*}, Mihaela Pruteanu^{1†*}, Michael Kuhn^{2*}, Georg Zeller², Anja Telzerow¹, Exene Erin Anderson¹, Ana Rita Brochado¹, Keith Conrad Fernandez¹, Hitomi Dose³, Hirotada Mori³, Kiran Raosaheb Patil², Peer Bork^{2,4,5,6} & Athanasios Typas^{1,2}

A few commonly used non-antibiotic drugs have recently been associated with changes in gut microbiome composition, but the extent of this phenomenon is unknown. Here, we screened more than 1,000 marketed drugs against 40 representative gut bacterial strains, and found that 24% of the drugs with human targets, including members of all therapeutic classes, inhibited the growth of at least one strain *in vitro*. Particular classes, such as the chemically diverse antipsychotics, were overrepresented in this group. The effects of human-targeted drugs on gut bacteria are reflected on their antibiotic-like side effects in humans and are concordant with existing human cohort studies. Susceptibility to antibiotics and human-targeted drugs correlates across bacterial species, suggesting common resistance mechanisms, which we verified for some drugs. The potential risk of non-antibiotics promoting antibiotic resistance warrants further exploration. Our results provide a resource for future research on drug-microbiome interactions, opening new paths for side effect control and drug repurposing, and broadening our view of antibiotic resistance.

Pharmaceutical agents have both beneficial and undesirable effects. Studies on the mechanisms of action and off-target spectra of various drugs aim to improve their efficacy and reduce their side effects. Although many drugs have gastrointestinal side effects and the gut microbiome itself is pivotal for human health¹, the role of the gut microbiota in these processes is rarely considered. Recently, consumption of drugs designed to target human cells and not microbes, such as anti-diabetics (metformin²), proton pump inhibitors (PPIs)^{3,4}, nonsteroidal anti-inflammatory drugs⁵ and atypical antipsychotics (AAPs)⁶, has been associated with changes in microbiome composition. A larger cohort study suggested that medication can alter gut microbiome composition more generally⁷. As it is unclear whether such effects are direct and go beyond the few drug classes studied, we systematically profiled interactions between drugs and individual gut bacteria. We aimed to generate a comprehensive resource of drug actions on the microbiome, which could facilitate more in-depth clinical and mechanistic studies, ultimately improving therapy and drug design.

A high-throughput drug screen on gut bacteria

To systematically map interactions between drugs and human gut bacteria, we monitored the growth of 40 representative isolates upon treatment with 1,197 compounds in modified Gifu anaerobic medium (mGAM) broth, which partially recapitulates the species relative abundances in human gut microbiomes⁸, under anaerobic conditions at 37°C (Extended Data Fig. 1a). We used the Prestwick Chemical Library, which consists mostly of off-patent Federal Drug Administration (FDA)-approved compounds with high chemical and pharmacological diversity. Most compounds are administered to humans (1,079), and they cover all main therapeutic classes (Supplementary Table 1). Three quarters (835) of the compounds are human-targeted drugs (that is, have molecular targets in human cells), whereas the rest are anti-infectives: 156 with antibacterial activity (144 antibiotics, 12 anti-septics) and 88 effective against fungi, viruses or parasites (Fig. 1a).

All compounds were screened at 20 µM, which is within the range of what is commonly used in high-throughput drug screens⁹.

For our screen to be representative of the gut microbiome of healthy individuals, we selected a set of ubiquitous gut bacterial species (Supplementary Table 2). Prevalence and abundance in the human gut, and phylogenetic diversity, were our main selection criteria (Extended Data Fig. 1b), although we were occasionally constrained by strain unavailability or irreproducible growth in mGAM. In total, we included 40 human gut isolates from 38 bacterial species and 21 genera (*Escherichia coli* and *Bacteroides fragilis* were represented by two strains each), accounting together for 78% of the median assignable relative abundance of the human gut microbiome at genus level (60% at species level; Extended Data Fig. 1c). Most strains were commensals, covering 31 of 60 sequenced species detected at a relative abundance of 1% or more and prevalence of at least 50% in fecal samples from asymptomatic humans from three continents (Extended Data Fig. 1d). In addition, the set included four pathobionts (*Clostridium difficile*, *Clostridium perfringens*, *Fusobacterium nucleatum* and an enterotoxigenic strain of *B. fragilis*), a probiotic (*Lactobacillus paracasei*) and two commensal *Clostridia* species (*C. ramosum* and *C. saccharolyticum*). All 38 species are found in the gut of healthy individuals and are part of a larger strain resource panel for the healthy human gut microbiome⁸.

We screened all compounds in multiwell plates, measuring optical density over time to monitor growth, and quantifying the area under the growth curve (AUC) up to the time point at which controls with unperturbed growth transitioned to stationary phase (see Methods; Extended Data Fig. 2). We obtained at least three biological replicates per strain, and these replicates correlated highly (Extended Data Fig. 2c). We then tested for significant deviations from the normalized AUC distribution of samples with unperturbed growth, combining *P* values across replicates and correcting for multiple hypothesis testing on the complete matrix of compounds and strains (see Methods; Extended Data Fig. 2). Drugs that significantly reduced the growth of

¹European Molecular Biology Laboratory, Genome Biology Unit, 69117 Heidelberg, Germany. ²European Molecular Biology Laboratory, Structural and Computational Biology Unit, 69117 Heidelberg, Germany. ³Graduate School of Biological Sciences, Nara Institute of Science and Technology, 630-0101 Ikoma, Japan. ⁴Max-Delbrück-Centre for Molecular Medicine, 13125 Berlin, Germany. ⁵Molecular Medicine Partnership Unit, 69120 Heidelberg, Germany. ⁶Department of Bioinformatics, Biocenter, University of Würzburg, 97024 Würzburg, Germany. [†]Present address: Institute for Biology, Humboldt University Berlin, 10115 Berlin, Germany.

*These authors contributed equally to this work.

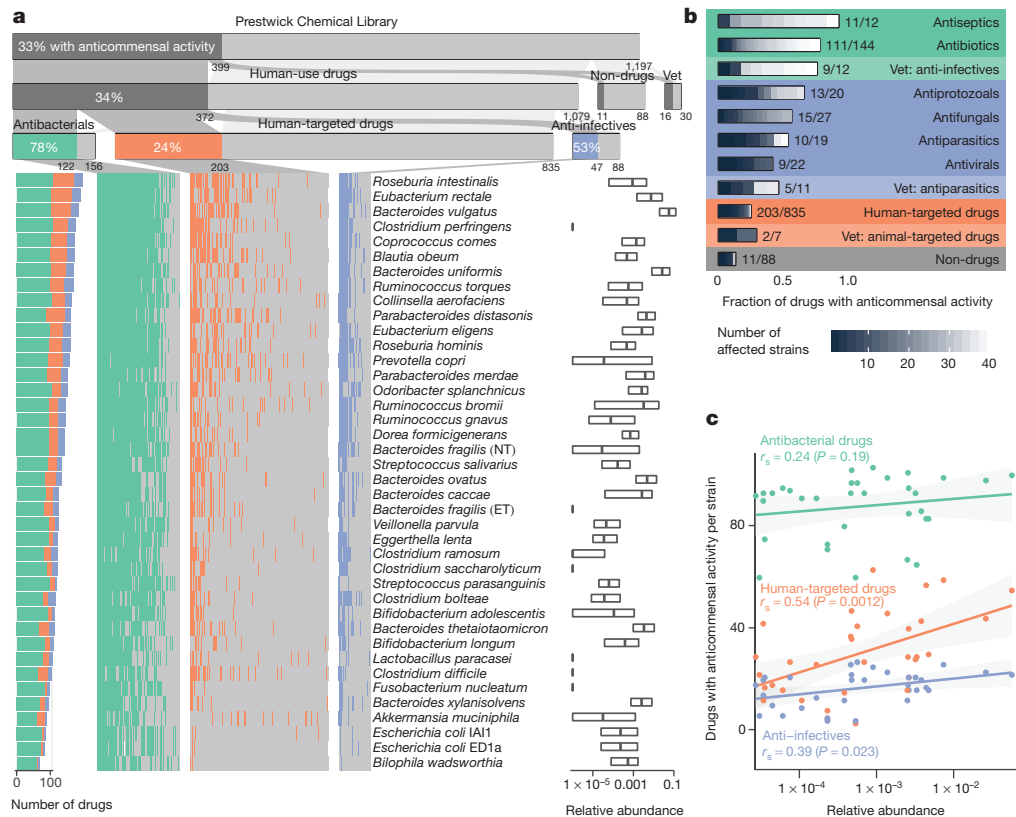


Figure 1 | Systematic profiling of marketed drugs on a representative panel of human gut microbial species. a, Broad impact of pharmaceuticals on the human gut microbiota. Compounds from the Prestwick Chemical Library are divided into drugs used in humans, drugs used exclusively in animals (vet) and compounds without medical or veterinary use (non-drugs). Human-use drugs are further categorized according to targeted organism. Strain–drug pairs (that is, instances in which a drug significantly reduced the growth of a specific strain; see Methods) are highlighted with a vertical coloured bar in the matrix. Bacterial strains are sorted by drug sensitivity. The relative abundances of each strain in four cohort studies of healthy individuals are displayed on

at least one strain (false discovery rate (FDR) < 0.01), were classified as hits with anticommonsensal activity (Supplementary Table 3a), reflecting their potential to modulate the human gut microbiota.

Of the 156 antibacterials tested, 78% were active against at least one species, typically with a broad activity spectrum (Fig. 1a, b). Inactive antibiotics belong mainly to the sulfonamides (which are inactive in our medium according to the manufacturer's guidelines), aminoglycosides (which have compromised activity under anaerobic conditions¹⁰) and specific antimycobacterial drugs. Antibiotics are used to inhibit pathogens, but as expected, also target gut commensals. The medical importance of this collateral damage to the resident microbiome has recently been becoming clearer¹¹. Nevertheless, to our knowledge, drug–microbiome species relationships have not previously been mapped at this scale.

Notably, 27% of the non-antibiotic drugs were also active in our screen. More than half of the anti-infectives against viruses or eukaryotes exhibited anticommonsensal activity (47 drugs; Fig. 1a, b). Antibacterial activity has been previously reported for many of these drugs, including the antifungal imidazoles¹² (11 in our screen), but not for others (for example, the antivirals saquinavir and trifluridine). More noteworthy is the anticommonsensal activity of 203 (24%) of the human-targeted drugs. Most were effective against only a few strains, with the exception of 40 drugs that affected at least 10 strains. Fourteen of these had, to our knowledge, not been previously reported to have direct antibacterial activity (Supplementary Table 3b). Among known human-targeted drugs with anticommonsensal activity, auranofin has

the right (boxes correspond to interquartile range (IQR) and central line to median relative abundance). **b**, Fractions of drugs with anticommonsensal activity by sub-category. Grey scale within bars denotes inhibition spectrum (the number of affected strains per drug). **c**, Correlation between species abundance in the human microbiome and drug sensitivity. For each strain ($n = 40$), the number of drugs that affect its growth is plotted against its median relative abundance in the human gut microbiome. Lines depict the best linear fit, r_s the Spearman correlation and grey shading the 95% confidence interval of the linear fit. All drugs, and in particular human-targeted drugs, inhibit the growth of more abundant species more than that of less abundant species.

recently been reported to have broad-spectrum bactericidal activity¹³, and the ovulation stimulant clomiphene inhibits a conserved bacterial enzyme in the synthesis of an essential precursor for cell wall carbohydrate polymers¹⁴. Such drugs or their scaffolds can be used for repurposing towards broad-spectrum antibiotics, especially as many have minimal inhibitory concentrations (MICs) in the sub-microgram per millilitre range (Supplementary Table 4). By contrast, the microbial narrow-spectrum specificity of most human-targeted drugs could aid the development of microbiome modulators.

Bacterial species showed varied responses to drugs, with the abundant *Roseburia intestinalis*, *Eubacterium rectale* and *Bacteroides vulgatus* being the most sensitive, and γ -proteobacteria representatives being the most resistant (Fig. 1a). Overall, species with higher relative abundance across healthy individuals were significantly more susceptible to human-targeted drugs ($P = 0.0012$ based on Spearman correlation; Fig. 1c). This suggests that human-targeted drugs have an even larger impact on the gut microbiome, with key species related to healthy status¹⁵, such as major butyrate producers (*E. rectale*, *R. intestinalis*, *Coprococcus comes*) and propionate producers (*B. vulgatus*, *Prevotella copri*, *Blautia obeum*)¹⁶, and enterotype drivers (*P. copri*)¹⁷, being relatively more affected.

Dose relevance and validation of the drug screen

We sought to address how close the screening concentration (20 μ M) was to drug concentrations in the terminal ileum and colon, where most gut microbes reside¹⁸. However, drug concentrations are

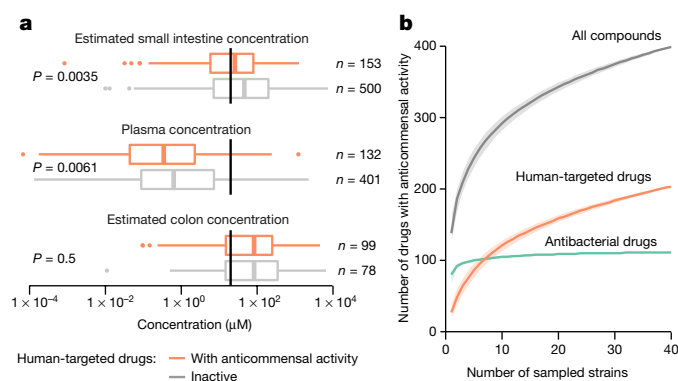


Figure 2 | Evaluating human-targeted drugs with anticomensal activity. **a**, Estimated small intestine and colon concentrations and measured plasma concentrations of human-targeted drugs with (orange) and without (grey) anticomensal activity in our screen (see Methods; Extended Data Fig. 3). For both active and inactive compounds, the median estimated small intestine and colon concentrations are higher than the screened concentration ($20 \mu\text{M}$, black vertical lines), whereas plasma concentrations are lower. Non-hits in our screen generally reached higher plasma and small intestine concentrations (two-sided Wilcoxon rank sum test). Box plots: centre line, median; limits, upper and lower quartiles; whiskers, $1.5 \times \text{IQR}$; points, outliers. **b**, Rarefaction analysis indicates that anticomensal activity would be discovered for more human-targeted drugs if we screened additional strains.

systematically measured only in blood; there, human-targeted drugs have on average an order of magnitude lower concentrations than in our screen (Fig. 2a, Extended Data Fig. 3). We deduced colon concentrations on the basis of drug excretion patterns from published work, and small intestine concentrations on the basis of daily doses of individual drugs (Supplementary Table 1) and a measured example of duodenal concentrations for the well-absorbed drug posaconazole¹⁹ (see Methods). Based on these approximations, $20 \mu\text{M}$ was below the median small intestine and colon concentration of the human-targeted drugs tested here (Fig. 2a, Extended Data Fig. 3). Notably, human-targeted drugs that showed anticomensal activity had lower plasma and estimated small intestinal concentrations than ones with no such activity (Fig. 2a; $P = 0.0061$ and $P = 0.0035$, respectively, two-sided Wilcoxon rank sum test; we have fewer colon concentration estimates owing to data availability), suggesting that more human-targeted drugs would inhibit bacterial growth if probed at higher doses, closer to physiological concentrations. A case in point is metformin, which was recently identified as the key contributor to changes in the human gut microbiome composition of patients with type II diabetes², but lacked anticomensal activity in our screen. Metformin reaches $10\text{--}40 \mu\text{M}$ in the plasma of treated patients with type II diabetes, but its small intestine concentration is $30\text{--}300$ -fold higher²⁰, which matches our estimates of small intestine and colon concentrations (1.5 mM). When we probed for higher, more physiological intestinal metformin concentrations, 3 of 22 tested strains were inhibited at concentrations below 1.5 mM (Extended Data Fig. 4a).

We also benchmarked our screen with an independent set of experiments, measuring IC_{25} (the drug concentration conferring 25% growth inhibition) for 25 selected drugs in a subset of up to 27 strains (see Methods). This analysis revealed excellent precision (94%), but slightly lower recall (85%) (Extended Data Fig. 5a, b). False negatives, that is, drugs with anticomensal activity missed in our screen, were due to specific chemicals that probably lost activity during screening (Extended Data Fig. 5d), and our stringent FDR cutoff for calling hits. Increasing this cutoff to 0.1 would almost double the fraction of drugs with anticomensal activity (Extended Data Fig. 5c). In addition, we found that more species were inhibited at higher concentrations (Extended Data Fig. 5d, Supplementary Table 4), and that IC_{25} values

were mainly below the estimated gut concentrations and occasionally below plasma concentrations (Extended Data Fig. 6).

Furthermore, we screened only a representative subset of species, but the gut microbiome of an individual harbours hundreds of species and an even larger strain diversity²¹. Rarefaction analysis indicates that if more gut species were tested, the fraction of human-targeted drugs with anticomensal activity would increase (Fig. 2b).

In summary, we probed human-targeted drugs largely within physiologically relevant concentrations and our data are likely to under-report the impact of human-targeted drugs on gut bacteria.

Concordance with patient data

Having demonstrated that many human-targeted drugs inhibit gut bacteria *in vitro* at relevant doses, we searched for evidence that such effects manifest *in vivo* in the human gut. We reviewed all available clinical cohort data from metagenomics association studies and compared it to our screen if studies had enough statistical power and affected taxa that overlapped with those tested here. We found suitable studies for PPIs, AAPs, and seven further drugs, spanning altogether five different drug classes according to Anatomical Therapeutic Chemical (ATC) classification. All three PPI representatives in our screen exhibited broad anticomensal activity, similar to the microbiome changes that have been reported in patients taking PPIs^{3,4} (Fig. 3a): taxa with reduced abundance in patients exhibited reduced growth in our screen and taxa enriched in patients were rarely inhibited by PPIs *in vitro* (Extended Data Fig. 7a). This suggests that PPIs directly influence the gut microbiome composition, in addition to changing the stomach pH and thus affecting which bacteria reach the gut^{3,4}. Concordance was similarly high for many microbe–drug associations identified in a large Flemish cohort⁷ for the immunosuppressive agent azathioprine, the antidepressant venlafaxine, the anti-inflammatory mesalazine, aminosalicilate, progesterone, oestrogens and amoxicillin; the only exception was another antibiotic, nitrofurantoin (Extended Data Fig. 7b, c). We also compared our data to a study that reported a reduction in *Akkermansia* levels in the gut of patients treated with AAPs⁶. Our screen included six of the ten AAPs investigated in that study. We found that *Akkermansia muciniphila* was more sensitive than other strains to these AAPs ($P = 0.09$; two-sided Wilcoxon rank sum test), while being more resistant to other human-targeted drugs ($P = 0.0005$, two-sided Wilcoxon rank sum test; Extended Data Fig. 7d). Finally, we found high concordance between a longitudinal microbiome study of patients taking metformin and our IC_{25} data for the same drug (Extended Data Fig. 4b).

Metagenomics association studies and our *in vitro* study have distinct limitations. We screened a subset of species, mostly one strain per species, out of the context of microbial communities and the host. Cohort studies can be underpowered or biased by methodological approaches and confounding factors, and may detect indirect effects. Nonetheless, we find high concordance between the effects of drugs *in vitro* and in humans, confirming clinical relevance and direct anticomensal activity for the aforementioned cases.

To assess the physiological relevance of our screen further, we investigated the registered side effects of these drugs in humans. We first identified side effects enriched in antibiotics for systemic use compared to those found in all other drugs in the SIDER database²². We identified 69 side effects that were enriched in antibiotics (see Methods; Supplementary Table 5). These antibiotic-related side effects occurred more often in clinical trials of human-targeted drugs with anticomensal activity than in trials of compounds that were inactive in our screen ($P = 0.002$, two-sided Wilcoxon rank sum test), whereas no significant difference was observed for placebo-treated patients (Fig. 3b). This suggests that the collateral damage of human-targeted drugs on gut bacteria can be detected by higher occurrences of antibiotic-like side effects in patients.

We then tested whether this side effect signature predicted anticomensal activity of human-targeted drugs, which we could have

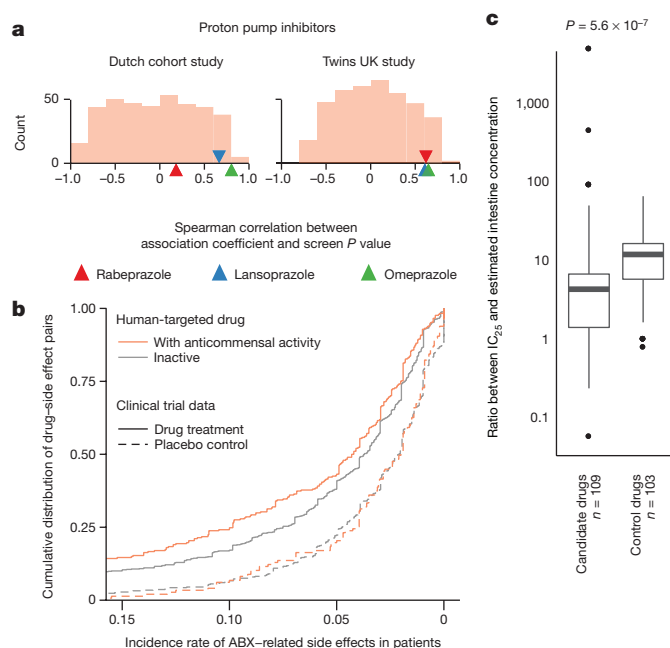


Figure 3 | Anticommensal activity of human-targeted drugs *in vitro* reflects patient data. **a**, Changes in microbiome composition of patients taking PPIs are consistent with drug effects in our screen. Displayed are Spearman correlation coefficients between *in vitro* growth inhibition *P* values and changes in taxonomic relative abundance after PPI consumption for corresponding taxa from two studies (Twins UK⁴ and Dutch³ cohorts; 229 of 1,827 and 211 of 1,815 individuals had taken PPIs, respectively). The histogram represents the background distribution of correlations between the *in vitro* data for all human-targeted drugs and the *in vivo* response to PPIs; correlations with PPIs are highlighted by triangles. **b**, Human-targeted drugs with anticommensal activity in our screen had a significantly higher incidence of antibiotic-related side effects (orange trace shows cumulative distribution, $n = 285$ drug-side effect pairs) in clinical trials compared to drugs without activity (grey trace, $n = 767$; $P = 0.002$, two-sided Wilcoxon rank sum test). Dashed lines indicate the incidence of the same side effects upon placebo treatment, with no significant difference between active ($n = 138$) and inactive drugs ($n = 474$). **c**, Based on similarity to antibiotic-related side effects, we selected 26 candidate and 16 control drugs for testing for anticommensal activity. Although both candidate and control drugs inhibited bacterial growth at higher concentrations, candidate drugs had anticommensal activity at significantly lower doses than control drugs after normalizing for estimated intestine concentrations ($P = 5.6 \times 10^{-7}$, two-sided Wilcoxon rank sum test). Box plots as in Fig. 2a, n denotes number of drug-strain pairs.

missed owing to the low drug concentration we used. We screened 26 candidate compounds that showed enrichment of antibiotic-related side effects and 16 that did not (control compounds) for effects on the growth of 18 bacterial strains (Extended Data Fig. 8), in concentrations up to 2.5 mM (Methods). Twenty-eight of these forty-two compounds inhibited the growth of at least one strain (Extended Data Fig. 8a–d), with both the fraction of active compounds and the number of affected strains being similar for both candidate and control compounds. However, when we normalized the measured IC_{25} by the estimated intestine concentration (based on the recommended single drug dose) to make amounts comparable between drugs, a significant difference was evident. Drugs that were predicted to be active had a median IC_{25} across all drug-strain pairs that corresponded to 4.3 drug doses, compared to 12 for control drugs ($P = 5.6 \times 10^{-7}$, two-sided Wilcoxon rank sum test; Fig. 3c). The IC_{25} corresponds to less than two drug doses in 34% of drugs with predicted activity, compared to just 8% for control drugs. Similarly, the IC_{25} is below the estimated colon concentration for 16/52 (31%) of candidate drug-strain pairs and for only 5/50 (10%) of control drug-strain pairs (Extended Data Fig. 8e).

In conclusion, human-targeted drugs with anticommensal activity have antibiotic-like side effects in humans, and for the few studies available, consumption of these drugs led to changes in taxa we also detected to be inhibited *in vitro*, implying that more drugs with anticommensal activity reported here will have an impact *in vivo*.

Features of drugs with anticommensal activity

Drugs from all major ATC indication areas exhibited anticommensal activity, with antineoplastics, hormones and compounds that target the nervous system inhibiting gut bacteria more than other medications (Extended Data Figs 9a, 10). Three ATC subclasses (antimetabolites, antipsychotics and calcium-channel blockers) were significantly enriched in hits (Extended Data Fig. 9a). Antimetabolites are used as chemotherapeutic and immunosuppressant agents, with their incorporation into RNA or DNA, or their interaction with synthesis enzymes being cytotoxic to human cells. Their molecular targets are often conserved in bacteria²³, explaining the observed effects and raising the possibility that antibacterial effects may also be directly involved in the development of mucositis during chemotherapy²⁴.

The enrichment in antipsychotics is intriguing, given that they target dopamine and serotonin receptors in the brain, which are absent in bacteria. Although phenothiazines are known to have antibacterial effects²⁵, nearly all subclasses of the chemically diverse antipsychotics exhibited anticommensal activity (Extended Data Fig. 9b). These drugs targeted a significantly more similar pattern of species than expected from their chemical similarity ($P = 2 \times 10^{-19}$, permutation test; Extended Data Fig. 9c). This raises the possibility that direct bacterial inhibition may not only manifest as side effect of antipsychotics²⁶, but also be part of their mechanism of action.

As different ATC indication areas contain chemically similar drugs, we investigated whether the chemical properties of drugs can influence their anticommensal activity (Extended Data Fig. 11a). To some degree, chemically similar human-targeted drugs had more similar effects in the screen than less similar drugs (Extended Data Fig. 11b). We tested several compound properties, including complexity, molecular weight, topological polar surface area (TPSA), volume and hydrophobicity (XLogP). Complex, heavier and larger compounds preferentially targeted Gram-positive bacteria, whereas Gram-negative bacteria were protected against such bulkier drugs by their selective outer membrane barrier (Extended Data Fig. 12). Owing to the vast number of chemical moieties present in drugs with anticommensal activity, we did not attempt an exhaustive enrichment analysis. Nevertheless, we did observe reactive nitro-groups being enriched in drugs with anticommensal activity ($P = 6.4 \times 10^{-6}$, Fisher's exact test), indicating that local chemical properties may confer antibacterial activity.

Human-targeted drugs may boost antibiotic resistance

There is a strong correlation between resistance to antibacterials and resistance to human-targeted drugs in our data that cannot be explained simply by general cell envelope composition, as there is no clear division between Gram-positive and Gram-negative bacteria (Fig. 4a). We reasoned that more specific but common mechanisms could confer resistance to both drug groups. To test this hypothesis, we selected TolC, known to efflux several antibiotics in *E. coli* and other bacteria²⁷, as a prominent representative of a general resistance mechanism against antibiotics. We profiled an *E. coli* $\Delta tolC$ mutant and its parental wild type (BW25113) against the Prestwick Chemical Library. *E. coli* lacking TolC not only became more sensitive to antibacterials (22 hits more than wild type), but also became equally more sensitive to human-targeted drugs (19 additional hits; Fig. 4a, Supplementary Table 6). This effect is not specific to *E. coli* or TolC, as a more antibiotic-resistant *B. uniformis* strain (HM-715) was also equally more resistant to human-targeted drugs (Fig. 4a).

While our data support a strong role for common general resistance mechanisms, there are also outliers to this trend, the most prominent being *C. difficile* and *P. distansoni* (Fig. 4a). For both, strong

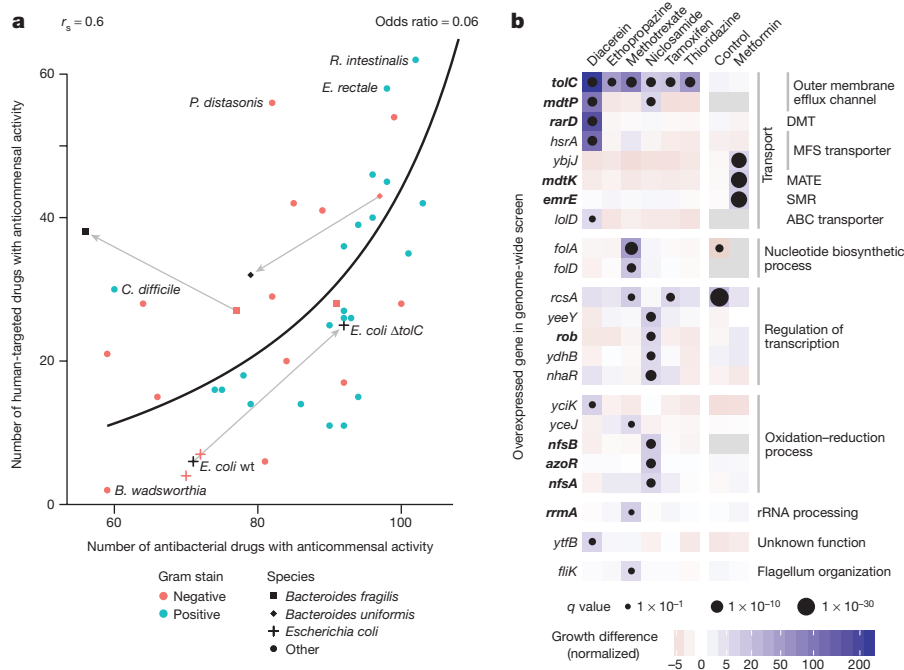


Figure 4 | Antibiotic resistance mechanisms protect against human-targeted drugs. **a**, Susceptibility to antibacterial agents and human-targeted drugs correlates across the 40 tested strains (Spearman correlation, $r_s = 0.6$ and a line depicting the nonlinear least-squares estimate of the odds ratio, $OR = 0.06$), suggesting common resistance mechanisms against both drug types. Knockout of a major antibiotic efflux pump (*tolC*) in the laboratory *E. coli* strain BW25113 (which behaves like the other two commensal *E. coli* strains in the screen) makes *E. coli* equally more sensitive to both antibacterials and human-targeted drugs. Two antibiotic-resistant isolates of *B. fragilis* (black square, HM-20) and *B. uniformis* (black diamond, HM-715) were screened in addition to the main screen, with only the latter showing a similar increase in resistance towards human-targeted drugs. **b**, Chemical genetic screen of an *E. coli*

antibiotic resistance²⁸ contrasted with relatively weaker resistance to human-targeted drugs. Similarly, an antibiotic-resistant *B. fragilis* strain, HM-20, was not equally resistant against human targeted drugs (Fig. 4a). These examples make the important distinction between specific antibiotic resistance mechanisms, which are irrelevant for resistance to human-targeted drugs, and more predominant, general mechanisms, which confer resistance to both drug groups.

To elucidate mechanisms conferring resistance against human-targeted drugs more systematically, we used a chemical genetic approach²⁹ and screened a genome-wide overexpression library in *E. coli* against seven non-antibiotics (six human-targeted drugs and niclosamide, an antiparasitic) that showed broad antimicrobial activity in our screen. As wild-type *E. coli* was one of the most resistant gut species (Fig. 4a), we used the $\Delta tolC$ mutant, which is sensitive to most of these drugs, allowing us to probe further resistance mechanisms. For all tested drugs except metformin, overexpression of *tolC* rescued *E. coli* growth, as expected. Furthermore, we identified a number of diverse transporter families that contributed to resistance against these drugs (Fig. 4b). Many of them have previously been linked to antibiotic resistance^{30–33}. Resistance was also acquired by overexpression of transcription factors (for example, *rob*, which controls efflux pump expression³⁴), the ribosome maturation factor *rrmA*, which plays a role in resistance to the antibiotic viomycin³⁵, and detoxification mechanisms (nitroreductases modify nitro-containing antibiotics³⁶). For methotrexate, we validated the known primary target in bacteria (*E. coli* dihydrofolate reductase)³⁷, illustrating the potential of this approach to identify bacterial mechanism of action of human-targeted drugs²⁹.

All of these results point to an overlap between resistance mechanisms against antibiotics and against human-targeted drugs, implying a

genome-wide overexpression library in seven non-antibiotics; all screens except for metformin were performed in $\Delta tolC$ background to sensitize *E. coli* to these drugs. Genes that when overexpressed significantly improved the growth of *E. coli* in the presence of at least one of the drugs are shown here; genes in bold have been previously associated with antibiotic resistance. Among them are genes encoding for transporters from different families: DMT (drug metabolite transporter), MFS (major facilitator superfamily), MATE (multidrug and toxin extrusion), SMR (small multidrug resistance) and ABC (ATP-binding cassette). Growth is measured by colony size (median $n = 4$)⁴⁰, colour depicts the normalized size difference from the median growth of all strains in the drug (more than sixfold difference), and dot size the significance (FDR-corrected $P < 0.1$). Control denotes the growth of the library without drug.

hitherto unnoticed risk of acquiring antibiotic resistance by consuming non-antibiotic drugs.

Discussion

We report a systematic drug screen against a reference panel of human gut bacteria. Twenty-seven per cent of non-antibiotics (24% of human-targeted drugs) inhibited the growth of at least one species. As we demonstrated, this is likely to be an underestimate owing to stringent thresholds for calling hits and the limited selection of bacterial strains screened. Many of the direct *in vitro* effects described here may translate into microbiome shifts *in vivo*, because (i) we used concentrations within the range of what is estimated to be found in the human gut for many drugs; (ii) our observations agree with the few clinical microbiome studies for which medication has been recorded; and (iii) the side effects of antimicrobial drugs in humans resemble those of antibiotics. Thus, our results underscore the necessity of accounting for potential medication-related confounding effects in future microbiome disease association studies. Moreover, one could speculate that pharmaceuticals, used regularly in our times, may be contributing to a decrease in microbiome diversity in modern Western societies³⁸.

Although the antibacterial potential of human-targeted drugs has been profiled repeatedly in the quest for new antimicrobials, previous efforts have focused on pathogenic and often multi-drug-resistant (MDR) bacterial species^{9,13,14}. We demonstrate that some of these species or their commensal relatives are the most drug-resistant in our screen (for example, γ -proteobacteria: *Bifidobacterium wadsworthia* and *E. coli* were affected by 2 and 4–7 human targeted drugs, respectively), that many human-targeted drugs have species-specific effects, and that resistance mechanisms to antibiotics and human-targeted drugs

partially overlap (thus, MDR species may be more resistant to human drugs too). Together, these findings explain why previous efforts have failed to register how many human-targeted drugs can inhibit bacteria.

Many pharmaceuticals influence the human gut microbiota. As gut bacteria, in turn, can also modulate drug efficacy and toxicity³⁹, the emerging drug–microbe network could guide therapy and drug development. The resource described here opens up new avenues for translational applications in mitigating drug side effects, improving drug efficacy, repurposing of human-targeted drugs as antibacterials or microbiome modulators, and controlling antibiotic resistance (see Supplementary Discussion). However, before any translational application can be pursued, our *in vitro* findings need to be tested rigorously *in vivo* (in animal models, pharmacokinetic studies and clinical trials) and understood better mechanistically.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March 2017; accepted 8 February 2018.

Published online 19 March 2018.

- Kährström, C. T., Pariente, N. & Weiss, U. Intestinal microbiota in health and disease. *Nature* **535**, 47 (2016).
- Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).
- Imhann, F. *et al.* Proton pump inhibitors affect the gut microbiome. *Gut* **65**, 740–748 (2016).
- Jackson, M. A. *et al.* Proton pump inhibitors alter the composition of the gut microbiota. *Gut* **65**, 749–756 (2016).
- Rogers, M. A. & Aronoff, D. M. The influence of non-steroidal anti-inflammatory drugs on the gut microbiome. *Clin. Microbiol. Infect.* **22**, 171–179 (2016).
- Flowers, S. A., Evans, S. J., Ward, K. M., McInnis, M. G. & Ellingrod, V. L. Interaction between atypical antipsychotics and the gut microbiome in a bipolar disease cohort. *Pharmacotherapy* **37**, 261–267 (2017).
- Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
- Tramontano, M. *et al.* Nutritional preferences of the human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-018-0123-9> (2018).
- Ejim, L. *et al.* Combinations of antibiotics and nonantibiotic drugs enhance antimicrobial efficacy. *Nat. Chem. Biol.* **7**, 348–350 (2011).
- Taber, H. W., Mueller, J. P., Miller, P. F. & Arrow, A. S. Bacterial uptake of aminoglycoside antibiotics. *Microbiol. Rev.* **51**, 439–457 (1987).
- Blaser, M. J. Antibiotic use and its consequences for the normal microbiome. *Science* **352**, 544–545 (2016).
- Rani, N., Sharma, A. & Singh, R. Imidazoles as promising scaffolds for antibacterial activity: a review. *Mini Rev. Med. Chem.* **13**, 1812–1835 (2013).
- Harbut, M. B. *et al.* Auranofin exerts broad-spectrum bactericidal activities by targeting thiol-redox homeostasis. *Proc. Natl Acad. Sci. USA* **112**, 4453–4458 (2015).
- Farha, M. A. *et al.* Antagonism screen for inhibitors of bacterial cell wall biogenesis uncovers an inhibitor of undecaprenyl diphosphate synthase. *Proc. Natl Acad. Sci. USA* **112**, 11048–11053 (2015).
- Pasolli, E., Truong, D. T., Malik, F., Waldron, L. & Segata, N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLOS Comput. Biol.* **12**, e1004977 (2016).
- Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
- Hens, B., Brouwers, J., Corsetti, M. & Augustijns, P. Supersaturation and precipitation of posaconazole upon entry in the upper small intestine in humans. *J. Pharm. Sci.* **105**, 2677–2684 (2016).
- Bailey, C. J., Wilcock, C. & Scarpello, J. H. Metformin and the intestine. *Diabetologia* **51**, 1552–1553 (2008).
- Schloissnig, S. *et al.* Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50 (2013).
- Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–D1079 (2016).
- Bodet, C. A., Ill, J., Jorgensen, J. H. & Drutz, D. J. Antibacterial activities of antineoplastic agents. *Antimicrob. Agents Chemother.* **28**, 437–439 (1985).
- Stringer, A. M., Gibson, R. J., Bowen, J. M. & Keefe, D. M. Chemotherapy-induced modifications to gastrointestinal microflora: evidence and implications of change. *Curr. Drug Metab.* **10**, 79–83 (2009).
- Sharma, S. & Singh, A. Phenothiazines as anti-tubercular agents: mechanistic insights and clinical implications. *Expert Opin. Investig. Drugs* **20**, 1665–1676 (2011).
- Morgan, A. P. *et al.* The antipsychotic olanzapine interacts with the gut microbiome to cause weight gain in mouse. *PLoS One* **9**, e115225 (2014).
- Li, X. Z., Plésiat, P. & Nikaido, H. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin. Microbiol. Rev.* **28**, 337–418 (2015).
- Nagy, E., Urbán, E., Nord, C. E. & ESCMID Study Group on Antimicrobial Resistance in Anaerobic Bacteria. Antimicrobial susceptibility of *Bacteroides fragilis* group isolates in Europe: 20 years of experience. *Clin. Microbiol. Infect.* **17**, 371–379 (2011).
- Cacace, E., Kritikos, G. & Typas, A. Chemical genetics in drug discovery. *Curr. Opin. Syst. Biol.* **4**, 35–42 (2017).
- Morita, Y. *et al.* NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob. Agents Chemother.* **42**, 1778–1782 (1998).
- Sulavik, M. C. *et al.* Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* **45**, 1126–1136 (2001).
- Nichols, R. J. *et al.* Phenotypic landscape of a bacterial cell. *Cell* **144**, 143–156 (2011).
- Nasie, I., Steiner-Mordoch, S. & Schuldiner, S. New substrates on the block: clinically relevant resistances for EmrE and homologues. *J. Bacteriol.* **194**, 6766–6770 (2012).
- Ariza, R. R., Li, Z., Ringstad, N. & Demple, B. Activation of multiple antibiotic resistance and binding of stress-inducible promoters by *Escherichia coli* Rob protein. *J. Bacteriol.* **177**, 1655–1661 (1995).
- Gustafsson, C. & Persson, B. C. Identification of the *rrmA* gene encoding the 23S rRNA m1G745 methyltransferase in *Escherichia coli* and characterization of an m1G745-deficient mutant. *J. Bacteriol.* **180**, 359–365 (1998).
- Roldán, M. D., Pérez-Reinado, E., Castillo, F. & Moreno-Vivián, C. Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *FEMS Microbiol. Rev.* **32**, 474–500 (2008).
- Matthews, D. A. *et al.* Dihydrofolate reductase: x-ray structure of the binary complex with methotrexate. *Science* **197**, 452–455 (1977).
- Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183 (2015).
- Spanogiannopoulos, P., Bess, E. N., Carmody, R. N. & Turnbaugh, P. J. The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016).
- Kritikos, G. *et al.* A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nat. Microbiol.* **2**, 17014 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Beltrao (EBI), K. C. Huang (Stanford) and F. Cabreiro (UCL) for feedback on the manuscript; F. Rippmann (Merck KGaA) for pointing to the delayed onset of antipsychotics; S. Wicha (University of Hamburg) for discussions on drug concentrations; J. Overington (Medicines Discovery Catapult) for help with drug plasma concentrations, and members of all four laboratories for fruitful discussions (in particular T. Hodges for suggestions on the manuscript and M. Driessen for experimental support). We thank the EMBL mechanical workshop for the custom-made incubator. We acknowledge funding from EMBL and the Microbios grant (ERC-AdG-669830). L.M. and M.P. were supported by the EMBL Interdisciplinary Postdoc (EIPoD) programme under Marie Skłodowska Curie Actions COFUND (grant 291772). A.Te. and A.R.B. were supported by a Sofja Kovalevskaja Award of the Alexander von Humboldt Foundation to A.Ty.

Author Contributions The study was conceived by K.R.P., P.B. and A.Ty., designed by L.M., M.P., G.Z., A.R.B. and A.Ty., and supervised by K.R.P., P.B. and A.Ty. *In vitro* screening was established by M.P. and performed by L.M., M.P., A.Te. and K.C.F. Follow-up and validation experiments were conducted by L.M., M.P. and E.E.A. H.D. and H.M. constructed and provided the Transbac library. Data preprocessing was performed by M.K. and G.Z.; statistical analyses by M.K.; data curation by L.M., M.K. and E.E.A.; data interpretation by L.M., M.P., M.K., G.Z., K.R.P., P.B. and A.Ty. L.M., M.K., G.Z., K.R.P., P.B. and A.Ty. wrote the manuscript with input from M.P. and A.R.B.; L.M., M.K. and G.Z. designed figures with input from K.R.P., P.B. and A.Ty. All authors approved the final version for publication.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.Ty. (typas@embl.de), G.Z. (zeller@embl.de), K.R.P. (patil@embl.de) and P.B. (bork@embl.de).

Reviewer Information *Nature* thanks K. Lewis, H. B. Nielsen, G. Wright and R. Xavier for their contribution to the peer review of this work.

METHODS

Bacterial strains and growth conditions. Bacterial isolates used in this study were purchased from DSMZ, BEI Resources, ATCC and Dupont Health & Nutrition, or were gifts from the Denamur Laboratory (INSERM). All strains were recovered in their recommended rich medium (resource and literature). The screen and validation experiments were performed in mGAM (HyServe GmbH & Co. KG, Germany, produced by Nissui Pharmaceuticals)⁴¹, as almost all species could grow robustly in this medium in a manner that is reflective of their gut abundance⁸. Because we selected for robust growth, potential positive effects of drugs on growth could not be detected. Only one strain was grown in Todd-Hewitt Broth (Sigma-Aldrich), one in a 1:1 mixture of mGAM and gut microbiota medium⁴² and, for one strain, mGAM was supplemented with 60 mM sodium formate and 10 mM taurine (see also Supplementary Table 2). All media were pre-reduced at least 1 day before use under anoxic conditions in an anaerobic chamber (Coy Laboratory Products Inc.) (2% H₂, 12% CO₂, rest N₂) and all experiments were performed under anaerobic conditions at 37°C unless specified otherwise. No statistical methods were used to predetermine sample size.

Species selection. To select a representative core of species in the human gut microbiome, we analysed 364 fecal metagenomes from asymptomatic individuals from three continents^{43–46}. Species were defined and their abundance quantified as previously described^{47,48}. A core set of 60 microbiome species was defined (Extended Data Fig. 1b–d), and from this core, 31 species were selected for this screen. Seven additional species were selected for reasons explained in the main text.

Screen of the Prestwick Chemical Library. Preparation of screening plates. The Prestwick Chemical Library was purchased from Prestwick Chemical Inc. with compounds coming dissolved in dimethyl sulfoxide (DMSO) at a concentration of 10 mM. Compounds were re-arrayed to redistribute the DMSO control wells in each plate and to minimize the total number of 96- and 384-well plates (4 × 384-well plates or 14 × 96-well plates). At the same time, drugs were diluted to a concentration of 2 mM to facilitate further aliquoting, and these plates were stored at –30°C. For each experimental batch (10 replicates in 96-well plates; 20 replicates in 384-well plates), we prepared drug plates in the respective growth medium (2 × for 96-well plates, 1 × for 384-well plates), and stored them at –30°C until use (maximum 2 months). Before inoculation, plates were thawed and pre-reduced in the anaerobic chamber overnight. The Biomek FXP (Beckman Coulter) liquid handling system was used for all rearranging and aliquoting of the library compounds. **Inoculation.** Strains were grown twice overnight to make sure we had a robustly and uniformly growing culture before inoculating the screening plates. For 96-well plates, the second overnight culture was diluted to fresh medium in order to reach 2 × the desired starting optical density (OD) at 578 nm. Next, 50 µl of this diluted inoculum was added to wells containing 50 µl of 2 × concentrated drug in the respective culture medium using a multichannel pipetter. The final drug concentration was 20 µM and each well contained 1% DMSO. We inoculated 384-well plates with a 384 floating pin replicator VP384FP6S (V&P Scientific, Inc.), transferring 1 µl of appropriately diluted overnight culture to wells containing 50 µl of growth medium, 1% DMSO and 20 µM drug. For bacterial species that reached lower OD in overnight cultures we transferred twice 1 µl of appropriately adjusted OD culture. For both 96- and 384-well plates, the starting OD was 0.01 or 0.05, depending on the growth preference of the species (Supplementary Table 2).

Screening conditions. After inoculation, plates were sealed with breathable membranes (Breathe-Easy) to prevent evaporation and cross-contamination between wells, and incubated at 37°C without shaking. Growth curves were acquired by tracking OD at 578 nm with a microplate spectrophotometer (EON, Biotek). Measurements were taken every 1–3 h after 30–60 s of linear shaking, initially manually but later automatically using a microplate stacker (Biostack 4, Biotek), fitted inside a custom-made incubator (EMBL Mechanical Workshop). We collected measurements for 16–24 h. Each strain was screened in at least three biological replicates.

Normalization of growth curves and quantification of growth. Growth curves were analysed by plate. All growth curves within a plate were truncated at the time of transition from exponential to stationary phase. The end of the exponential phase was determined automatically by finding the peak OD (using the median across all compounds and control wells, and accounting for a small increase during the stationary phase) and verified by inspection. Using this time point allowed us to capture the effects of drugs on lag phase, growth rate and stationary phase plateau (Extended Data Fig. 2a). Time points with sudden spikes in OD (for example, caused by condensation) were removed, and growth curves were discarded completely if they had too many missing time points (Extended Data Fig. 2a). Similarly, growth curves were discarded if the OD fell far outside the normal range (for example, caused by coloured compounds). Three compounds had to be completely excluded from the analysis, as they caused aberrant growth curves: Chicago sky blue 6B, mitoxantrone and verteporfin.

Growth curves were processed by plate to set the median OD at the start and end time points to 0 and 1, respectively. Then we defined reference compounds

across all replicates as those that did not reduce growth significantly for most drugs, had measurements for >95% of all replicates, and for which the final OD was >0.5 for more than 142 out of 152 replicates. We used these reference compounds as representatives of uninhibited growth. As wells containing reference compounds outnumbered control wells within a plate, we used control wells only later to verify the *P* value calculation (Extended Data Fig. 2d). After identifying reference compounds, we rescaled growth curves such that the median growth of reference compounds at the end point was 1.

While growth curves in control wells and most wells with reference compounds followed the expected logistic growth pattern, a variety of deviations were observed for drugs that influenced growth. To quantify growth without relying on assumptions about the shape of the growth curve, we calculated the area under the curve (AUC) using the trapezoidal rule. Although we set the median starting OD to 0, the ODs of individual wells deviated from this. We used two methods to correct for this and determine the baseline for each growth curve (Extended Data Fig. 2a). First, a constant shift was assumed, subtracting the same shift from all time points of the growth curve such that the minimum is zero. Second, an initial perturbation was assumed that affects initial time points more than later time points (for example, condensation). To correct this, we first subtracted a constant shift as above, and then rescaled the curve such that a time point with an uncorrected OD of 1 also had an OD of 1 after correction. AUCs were calculated for both scenarios, rescaled such that the AUC of reference compounds was 1, and then for each compound the baseline correction that yielded an AUC closest to 1 (that is, normal growth) was selected.

AUCs are highly correlated to final ODs, with a Pearson correlation of 0.95 across all compounds and replicates. Nonetheless, we preferred to use AUCs to decrease the influence of the final time point, which will contain more noise than a metric based on all time points.

Identification of drugs with anticomensal activity. We detected hits from normalized AUC measurements using a statistical method that controls for multiple hypothesis testing and varying data quality. We fitted heavy-tailed distributions (scaled Student's *t*-distribution⁴⁹) to the wells containing reference compounds for each replicate and, separately, to each individual plate. These distributions captured the range of AUCs expected for compounds that did not reduce growth, and represented the null hypothesis that a given drug did not cause a growth defect in the given replicate or plate. We calculated one-sided *P* values from the cumulative distribution function of the fitted distribution. Within a replicate, each compound was associated with two *P* values: one from the plate on which it was measured, and one for the whole replicate. Of those two, the highest *P* value was chosen (conservative estimate) to control for plates with little or high noise, and varying levels of noise within the same replicate.

The resulting *P* values were well-calibrated (that is, the distribution of *P* values was close to uniform with the exception of a peak at low *P* values, Extended Data Fig. 2d) and captured the distribution of controls, which were not used for fitting the distribution and kept for validation. We then combined *P* values for a given drug and strain across replicates using Fisher's method. Lastly, we calculated the FDR using the Benjamini–Hochberg method⁵⁰ over the complete matrix of *P* values (1,197 compounds by 40 strains). After inspecting representative AUCs for compound–strain pairs at different FDR levels, we chose a conservative FDR cut-off of 0.01.

Drug indications, dose, and administration. We annotated drugs by their primary target organism on the basis of their WHO ATC classification, or, if there were uncertainties, based on manual annotation. Compounds were classified as: antibacterial drugs (antibiotics, antiseptics), anti-infective drugs (acting against protozoa, fungi, parasites or viruses), human-targeted drugs (that is, drugs whose mechanism of action affects human cells), veterinary drugs (used exclusively in animals), and finally non-drugs (which can be drug metabolites, drugs used only in research, or endogenous substances). If a human-use drug belonged to several classes, the drug class was picked according to this order of priority (from high to low): antibacterial, anti-infective, and human-targeted drug. This ensured that drugs used also as antibacterials were not classified in the other two categories.

Drugs from the Prestwick Chemical Library were matched against STITCH 4 identifiers⁵¹ using CART⁵². Identifiers that could not be mapped were annotated manually. Information about drug indications, dose and administration was extracted from the ATC classification system and Defined Daily Dose (DDD) database. Dose and administration data were also extracted from the Drugs@FDA resource. Doses that were given in grams were converted to mol using the molecular weight stated in the Prestwick library information files. When the dose guidelines mentioned salt forms, we manually substituted the molecular weight. Dose data from Drugs@FDA stated the amount of drug for a single dose (for example, a single tablet). Analysing the intersection between Drugs@FDA and DDD, we found that the median ratio between the single and daily doses was two. To combine the two data sets we therefore estimated the single dose as half of the daily dose (Supplementary Table 1).

In general, it is difficult to estimate effective drug concentrations in the intestine, as those depend on the dose, the speed of dissolution, uptake and metabolism by human cells and by bacteria, binding to proteins, and excretion mechanisms into the gut. To estimate gut concentrations of drugs based on their dose with a simple model, we relied on an *in situ* study for posaconazole¹⁹. When 40 mg (57 μ mol) of the drug is delivered to the stomach in either an acidic or a neutral solution, the maximum concentration in the duodenum reaches 26.3 ± 10.3 or $13.6 \pm 5.8 \mu$ M, respectively. This is equivalent to dissolving the drug in 300 ml (240 ml of water to swallow the pill as recommended for bioavailability/bioequivalence studies plus ~ 43 ml resting water in the small intestine⁵³) and an absorption rate of 90%. We collected doses for as many human-targeted drugs as we could find and used the above assumption to estimate small intestine concentrations. To estimate colon concentrations, we relied on reported fecal excretion data (Supplementary Table 1, gathered from DrugBank 5.0⁵⁴ and across the literature) assuming a single daily dose, 24 h transit time⁵⁵ and a volume of distribution in the colon of 0.6 l^{56} (Extended Data Fig. 3).

IC₂₅ determination and screen validation. To validate our screen, we selected 25 drugs including human-targeted drugs (19), antiprotzoals (3), one antiparasitic, one antiviral and one 'non-drug' compound. The human-targeted drugs spanned five therapeutic classes (ATC codes A, G, L, M and N). Our selection comprised mostly drugs with extended antibacterial activity in our screen (19 drugs hit >10 strains). This bias ensured that we could also evaluate false positives. We chose 15 strains to test IC₂₅s (that is the minimal concentration of drug that causes 25% growth inhibition), spanning different phyla (5) and including both sensitive (*E. rectale*, *R. intestinalis*) and resistant species (*E. coli* ED1a).

Compounds for validation were purchased from independent sources (Supplementary Table 1) and dissolved at $100\times$ starting concentration in DMSO. Twofold serial dilutions were prepared in 96-well U-bottom plates (as for the screen). Each row contained a different drug at eleven twofold dilutions and a control DMSO well in the middle of the row (in total eight drugs per plate). These master plates were diluted to $2\times$ assay concentration and 2% DMSO in mGAM (50 μ l) and stored at -30°C (<1 month). For the assay, plates were pre-reduced overnight in the anaerobic chamber, and mixed with an equal volume (50 μ l) of appropriately diluted overnight culture (prepared as described for screening section) to reach a starting OD₅₇₈ of 0.01 and a DMSO concentration of 1% across all wells. OD₅₇₈ was measured hourly for 24 h after 1 min of shaking. Experiments were performed in two biological replicates.

Growth curves were converted to AUCs as described above, using in-plate control wells (no drug) to define normal growth. For each concentration, we calculated the mean across the two replicates. We further enforced monotonicity to conservatively remove noise effects: if the AUC decreased for lower concentrations, it was set to the highest AUC measured at higher concentrations. The IC₂₅ was defined as the lowest concentration for which a mean AUC of below 0.75 was measured. In 68% of cases, IC₂₅s were equal between replicates and in a further 22%, there was a twofold change between replicates, which is within the twofold error margin reported for inhibitory concentrations⁵⁷. Additionally, MIC as listed in Supplementary Table 4 was defined as the lowest concentration for which the AUC dropped below 0.1. In the large-scale screen, we detected significant growth reductions, which do not necessarily correspond to complete growth inhibition (Extended Data Fig. 2b). To ensure comparability between the results of the validation procedure and the screen, we used the IC₂₅ metric for benchmarking. As inhibitory concentration calculations are known to have a twofold error margin⁵⁷, we considered an IC₂₅ of 10–40 μ M as being in agreement with the screening result (Extended Data Fig. 5a, b). A higher number of false negatives implies that more human-targeted drugs are likely to have anticomensal activity.

Analysis of side effects. Side effects of drugs were extracted from the SIDER 4.1 database²² using the mapping between Prestwick compounds and STITCH 4 identifiers described above. In SIDER, side effects are encoded using the MedDRA terminology, which contains lower-level terms and preferred terms. Of these, we used the preferred terms, which are more general. We excluded rare side effects that occurred for fewer than five drugs from the analysis. Drugs with fewer than seven associated side effects were discarded⁵⁸. In a first pass, we identified side effects associated with antibiotics in SIDER, by calculating for each side effect its enrichment for systemic antibiotics (ATC code J01) versus all other drugs using Fisher's exact test (P value cut-off: 0.05, correcting for multiple hypothesis testing using the Benjamini–Hochberg method). Antibiotics are typically administered in relatively high doses, and some of the enriched side effects might therefore be caused by a dose-dependent effect (for example, kidney toxicity). We therefore used an ANOVA (type II) to test whether the presence of side effects for a drug was more strongly associated with it being an antibiotic or with its (log-transformed) dose. Side effects that were more strongly associated with the dose were excluded from the list of antibiotic-related side effects.

Data on the incidence rates of side effects in patients was also extracted from SIDER 4.1. As different clinical trials can report different incidence rates, we computed the median incidence rate per drug–side effect pair. As SIDER also contains data on the incidence of side effects upon placebo treatment, we were able to ensure the absence of systematic biases.

Experimental validation of side effect-based predictions. Selected candidate and control compounds belonged to multiple therapeutic classes (ATC codes A, B, C, G, H, L, M, N, S for candidate compounds and A, C, D, G, H, M, N, R, S, V for control compounds). Compounds of interest were purchased from independent sources (Supplementary Table 1) and if possible, dissolved at 5 mM concentration in mGAM. Lower concentrations were used when the solubility limit was reached. Solutions were sterile filtered, and three fourfold serial dilutions were arranged in 96-well plates, aiming at covering a broad range of drug concentrations. Inoculation and growth curve acquisition was performed as described for the IC₂₅ determination experiments.

Chemical genetics in *E. coli*. *Conjugation of the TransBac overexpression plasmid library into *E. coli* Δ tolC.* The TransBac library, a new *E. coli* overexpression library based on a single-copy vector⁵⁹ (H.D. and H.M., unpublished resource) was conjugated in the BW25113 Δ tolC::Kan strain. The receiver strain (BW25113 Δ tolC::kan) was grown to stationary phase in LB medium, diluted to an OD₅₇₈ of 1, and 200 μ l was spread on an LB plate supplemented with 0.3 mM diaminopimelic acid (DAP). Plates were dried for 1 h at 37°C and then a 1536 colony array of the library carried within a donor strain (BW38029 Hfr (CIP8 *oriT::cat*) *dap*^{−60}) was pinned on top of the lawn. Conjugation was carried out at 37°C for ~ 6 h, and the first selection was done by pinning on LB plates supplemented with tetracycline only (10 μ g/ml) and growing overnight. Two more rounds of selection followed on LB plates containing both tetracycline (10 μ g/ml) and kanamycin (30 μ g/ml) to ensure killing of parental strains and select only for *tolC* mutants carrying the different plasmids.

Chemical genetic screen. The screen was carried out under aerobic conditions on solid LB Lennox medium (Difco), supplemented with 30 μ g/ml kanamycin, 10 μ g/ml tetracycline, the appropriate drug, and 0 or 100 μ M IPTG. Drugs were used at the following sub-inhibitory concentrations for the *tolC* mutant: diacerein 20 μ M, ethopropazine hydrochloride 160 μ M, tamoxifen citrate 20 μ M, niclosamide 1.25 μ M, thioridazine hydrochloride 40 μ M, methotrexate 320 μ M, or for the wild type: metformin 100 mM. The 1536 colony array of BW25113 Δ tolC::kan mutant carrying the TransBac collection was pinned on the drug-containing plates, and plates were incubated for 16–38 h at 37°C . In the case of metformin we used the version of the TransBac library in which each plasmid complements its corresponding barcoded single-gene deletion mutant⁵⁹, since we did not need to use the Δ tolC background to sensitize the cell. Growth of this library was determined at 0 and 100 mM metformin (both in the presence of 0, 50 and 100 μ M IPTG). All plates were imaged using an 18-megapixel Canon Rebel T3i and images were processed using the Iris software⁴⁰.

Data analysis. We used colony size to measure the fitness of the mutants on the plate. For standardization of colony sizes, we subtracted the median colony size and then divided by a robust estimate of the s.d. (removing outliers below the 1st and above the 99th percentile). We found edge effects affecting up to five rows and columns around the perimeter of the plate. We therefore first standardized colony sizes across the whole plate using only colony sizes from the inner part of the plate as reference. To remove the edge effects, we subtracted from each column its median colony size, and then from each row its median colony size. Finally, we standardized the adjusted colony sizes using the whole plate as reference. The distribution of adjusted colony sizes was right-skewed (that is, more outlier colonies with larger sizes), suggesting a log-normal distribution. At the same time, the presence of outliers suggested that a logarithmic equivalent of the Student's t -distribution with variable degree of freedom⁴⁹ would be more suitable. We fitted such a distribution for each plate and calculated P values for both tails of the distribution. This approach assumes that the overexpression of most genes does not affect growth in response to drug treatment. P values were combined using Fisher's method across replicates and IPTG concentrations (since we noticed that different IPTG concentrations resulted in largely the same results—that is, plasmids are leaky). We corrected for multiple hypothesis testing for each drug individually using the Benjamini–Hochberg method⁵⁰.

Analysis of common resistance mechanisms. To determine a relationship between the number of human-targeted drugs (h) and the number of antibacterial drugs (a) that affect each strain, we determined the odds ratio (OR):

$$\text{OR} = \frac{\frac{h}{H-h}}{\frac{a}{A-a}}$$

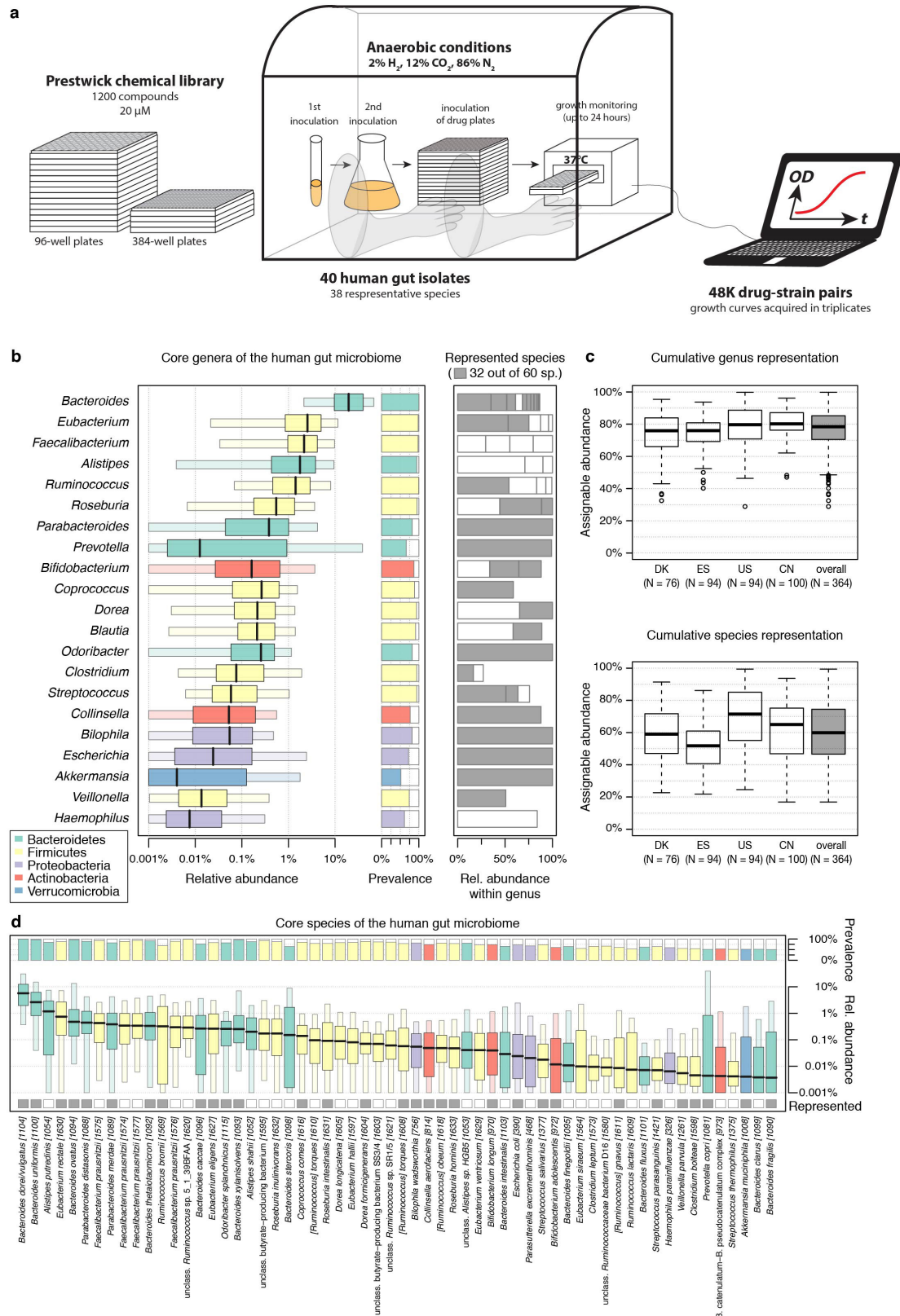
where $H = 203$ and $A = 122$ are the numbers of human-targeted and antibacterial drugs that show activity against any strain, respectively. We computed the nonlinear least-squares estimate for OR using the following equation:

$$\frac{h}{H-h} = \text{OR} \times \frac{a}{A-a}$$

Data availability. Data are available from FigShare: <http://dx.doi.org/10.6084/m9.figshare.4813882>. All data generated during this study are included in this published article and its Supplementary Information files.

Code availability. Scripts for analysing data and generating figures are available at https://git.embl.de/mkuhn/drug_impact_gut_bacteria. A snapshot of the repository has been deposited together with the data.

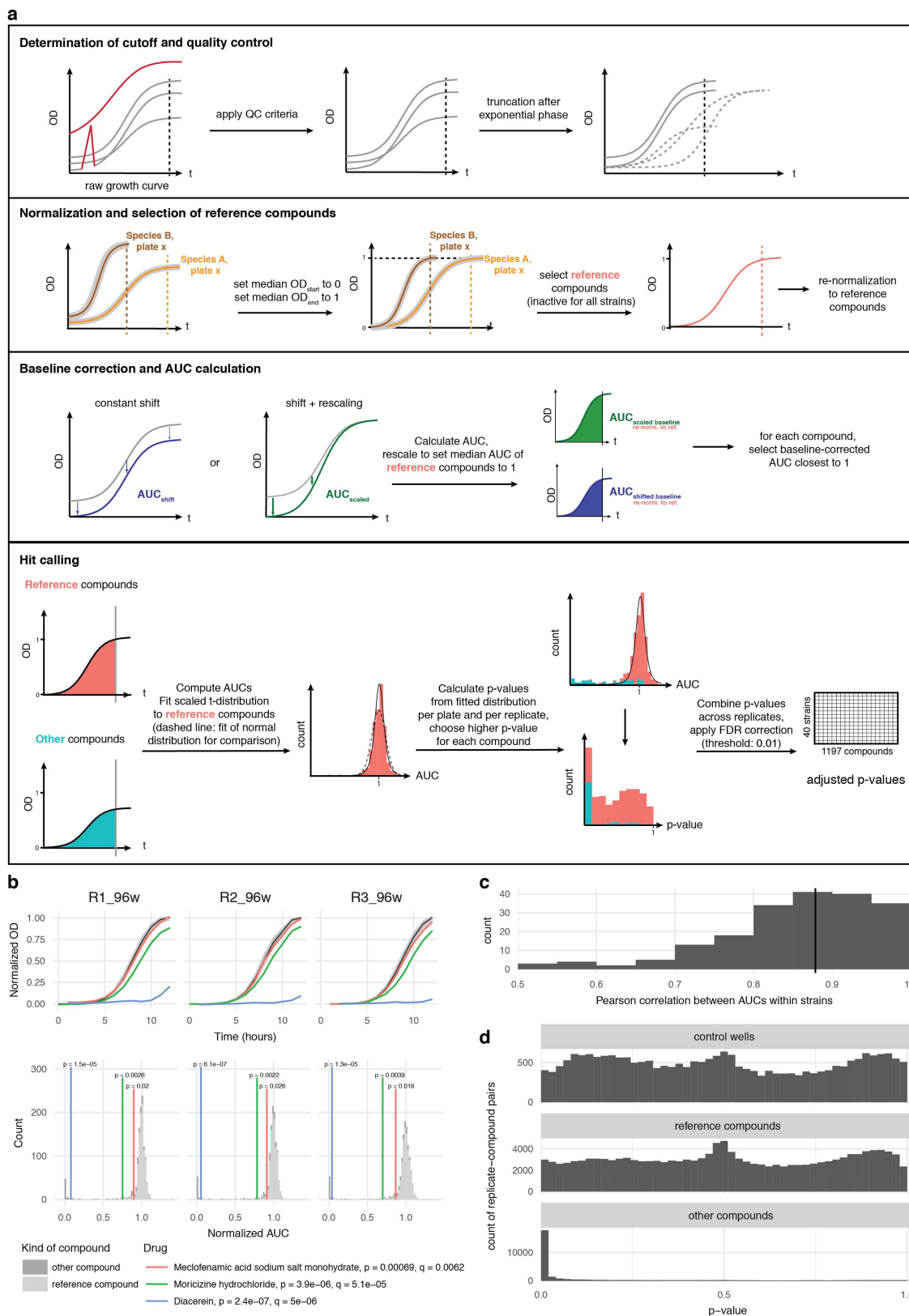
41. Rettedal, E. A., Gumpert, H. & Sommer, M. O. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nat. Commun.* **5**, 4714 (2014).
42. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).
43. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
44. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
45. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
46. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
47. Mende, D. R., Sunagawa, S., Zeller, G. & Bork, P. Accurate and universal delineation of prokaryotic species. *Nat. Methods* **10**, 881–884 (2013).
48. Kultima, J. R. *et al.* MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* **32**, 2520–2523 (2016).
49. Kruschke, J. K. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).
50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
51. Kuhn, M. *et al.* STITCH 4: integration of protein-chemical interactions with user data. *Nucleic Acids Res.* **42**, D401–D407 (2014).
52. Deghou, S. *et al.* CART-a chemical annotation retrieval toolkit. *Bioinformatics* **32**, 2869–2871 (2016).
53. Mudie, D. M. *et al.* Quantification of gastrointestinal liquid volumes and distribution following a 240 mL dose of water in the fasted state. *Mol. Pharm.* **11**, 3039–3047 (2014).
54. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
55. Kim, E. R. & Rhee, P. L. How to interpret a functional or motility test—colon transit study. *J. Neurogastroenterol. Motil.* **18**, 94–99 (2012).
56. Pritchard, S. E. *et al.* Fasting and postprandial volumes of the undisturbed colon: normal values and changes in diarrhea-predominant irritable bowel syndrome measured using serial MRI. *Neurogastroenterol. Motil.* **26**, 124–130 (2014).
57. Turnidge, J. & Paterson, D. L. Setting and revising antibacterial susceptibility breakpoints. *Clin. Microbiol. Rev.* **20**, 391–408 (2007).
58. Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J. & Bork, P. Drug target identification using side-effect similarity. *Science* **321**, 263–266 (2008).
59. Otsuka, Y. *et al.* GenoBase: comprehensive resource database of *Escherichia coli* K-12. *Nucleic Acids Res.* **43**, D606–D617 (2015).
60. Mori, H. *et al.* Identification of essential genes and synthetic lethal gene combinations in *Escherichia coli* K-12. *Methods Mol. Biol.* **1279**, 45–65 (2015).
61. Wu, H. *et al.* Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* **23**, 850–858 (2017).
62. Steinbeck, C. *et al.* Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des.* **12**, 2111–2120 (2006).
63. Kim, S. *et al.* PubChem substance and compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Screen set-up and species selection. **a**, Drugs from the Prestwick Chemical Library (arranged in either 96- or 384-well format) were diluted in growth medium (usually mGAM) and pre-reduced in a Coy anaerobic chamber before inoculation with one of forty different human gut microbes. Bacterial growth was monitored for 16–24 h at 37 °C. Growth curves were acquired at least in triplicate for each drug–microbe interaction (see Methods). **b**, Species with a minimum relative abundance of 1% in at least one sample and a prevalence of 50% across samples (the latter estimated by rarefying to 10,000 reads mapping to taxonomic markers) were included in the set of core species. Boxplots show relative abundances of core species grouped by genus (according to NCBI taxonomy) and coloured by phylum (see key). The inner box indicates the IQR, with the median as black vertical line; the outer bars extend to the 5th and 95th percentiles; circles, outliers. To the right of the boxplots, prevalence is depicted by bars, and next to this the species diversity is

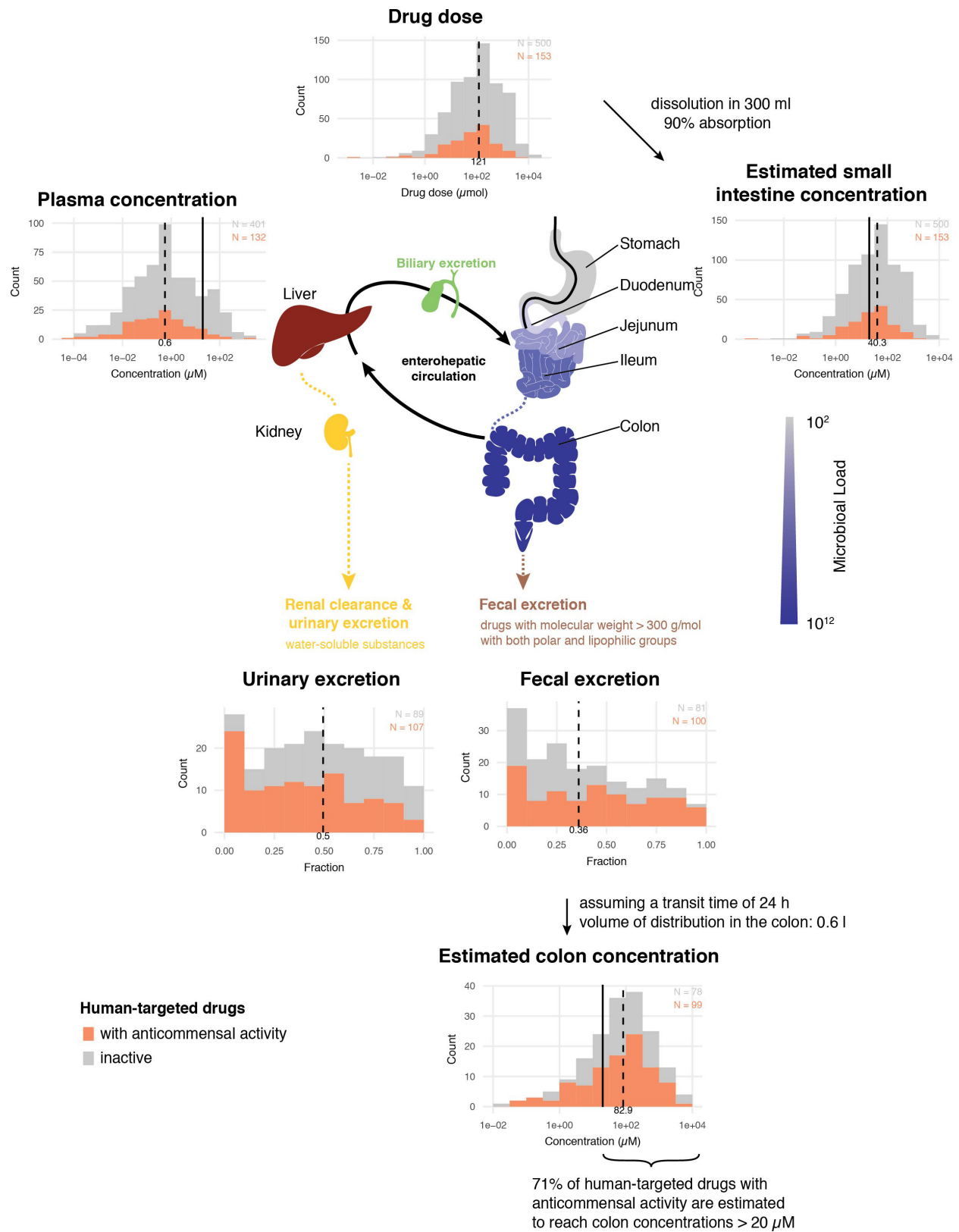
shown; grey boxes indicate species represented in the screen with box widths corresponding to mean relative abundance within the genus. **c**, Relative abundance of genera of which at least one species was represented in the screen cumulates to 78% of the assignable fraction of reads (median across all samples, upper panel); first four boxplots show abundance within each study identified by country codes underneath (DK: Denmark; ES: Spain; US: United States; CN: China)^{43–46}. When directly cumulating the relative abundance of represented species the corresponding median is 60% (lower panel). Boxes span the IQR and whiskers extend to the most extreme data points up to a maximum of 1.5 times the IQR. **d**, Core species are shown in the order of their median abundances across all samples. Relative abundance boxplots and prevalence bars are defined as in **b** and grey boxes underneath indicate species screened in this study. Numbers in brackets correspond to specI cluster identifiers (version 1)⁴⁷.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Data analysis pipeline for identifying compounds with anticomensal activity. **a**, Schematic overview of the data analysis pipeline. All steps (determination of time cutoff and removal of noisy points; normalization and selection of reference compounds; baseline correction and AUC calculation; and hit calling) are explained in detail in the Methods. On the first panel, dashed curves in the righthand plot depict the three possible effects that a drug can have on the growth of a microbe: increase the lag phase, decrease the growth rate or the stationary phase plateau. All effects are captured by cutting off the growth curves upon transition to stationary phase for most compounds (most drugs do not affect growth). On second panel, median growth rates for two drugs on same plate are depicted and normalized, whereas baseline correction (third panel) is applied at the individual wells. **b**, Growth curves (top, normalized OD) of *Bacteroides ovatus* in three exemplary drug cases

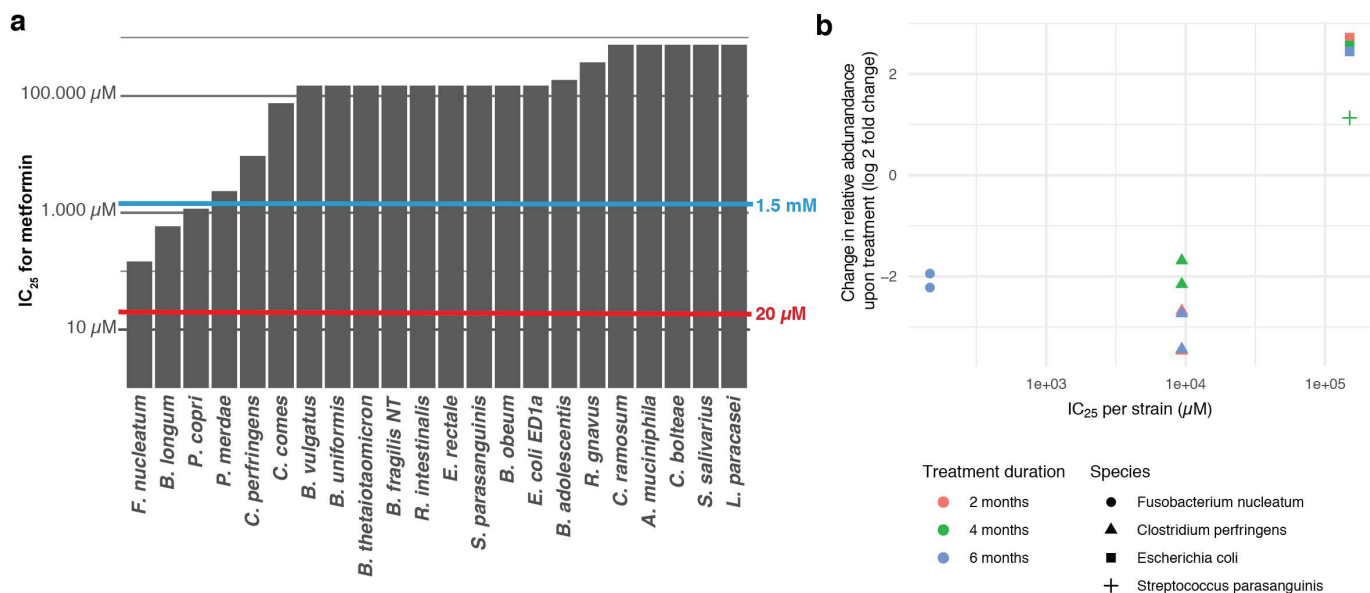
for the three biological replicates (meclofenamic acid (red), moricizine (green) and diacerein (blue)). Light and dark grey shades represent the 50% and 90% confidence intervals for normal growth, respectively. Bottom, normalized AUC histograms for all drugs in the three biological replicates for *B. ovatus*. Meclofenamic acid is just below the hit threshold, moricizine is a hit with partial but strong growth inhibition, and diacerein almost completely inhibits the growth of *B. ovatus*. **c**, For most species, correlation between replicates is very high (median: 0.88). **d**, For both controls and reference compounds, *P* values were approximately uniformly distributed. Determining the background distribution of uninhibited growth using reference compounds is validated by their very similar behaviour with control wells. Other drugs (that is, drugs not used as reference compounds) show clear enrichment of low *P* values.



Extended Data Figure 3 | See next page for caption.

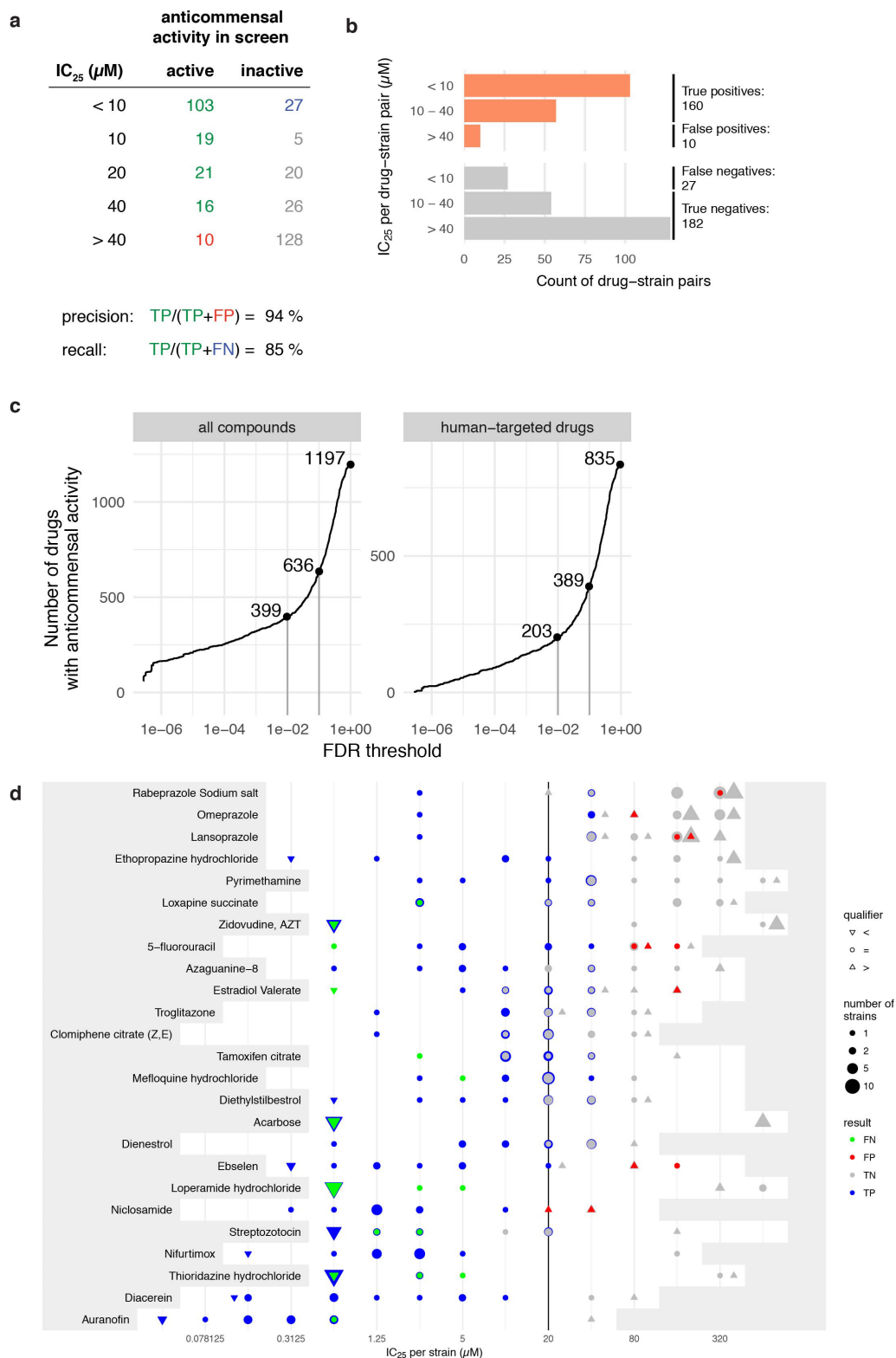
Extended Data Figure 3 | Anticommensal activity relative to compound- and compartment-specific drug concentrations. We made a simplified pharmacokinetic estimation of small intestine and colon concentrations by assuming that one dose of an orally administered drug (extracted from Drugs@FDA and Daily Defined Dose (DDD) of the ATC) reaches the intestine and is dissolved or absorbed similarly to the well-absorbed drug posaconazole¹⁹ (Supplementary Table 1). After absorption into the liver via the portal circulation, the drug enters circulation through the hepatic veins and reaches its characteristic plasma concentration. The two main routes of drug elimination are either secretion via kidneys and urine or secretion into the intestine via the biliary duct. In the intestine, drugs can be reabsorbed in a circuit called the enterohepatic cycle or excreted in stools. Compounds that are either poorly absorbed in the small intestine or secreted by bile reach the large intestine. Considering the measured excreted fraction of the drug in faeces (both changed and unchanged compound, as we do not know whether drug is metabolized

in liver or gut), and assuming a large intestinal transit time of 24 h⁵⁵ and a volume of distribution in the colon of 0.6 l⁵⁶, we estimated the colon concentrations of the human-targeted drugs in our screen (Supplementary Table 1). Histograms for drug dose, plasma concentration, estimated small intestine concentration, urinary and fecal excretion and estimated colon concentration depict the respective distributions for human-targeted drugs, colour coded according to their anticommensal behaviour in our screen. Dashed lines indicate medians and vertical lines highlight the drug concentration used in our screen (20 µM). Interactions between drugs and microbiota are possible throughout the entire gastrointestinal tract, with microbial load having a gradient-like distribution (ileum and colon containing the largest numbers); this can be disturbed during disease¹⁸. In addition, drugs can be modified at several stages: by host digestive and intestinal epithelial enzymes, by phase I and phase II metabolism in the liver and by microbial enzymes. Some of these processes neutralize each other, resulting in reconversion into the original compound.



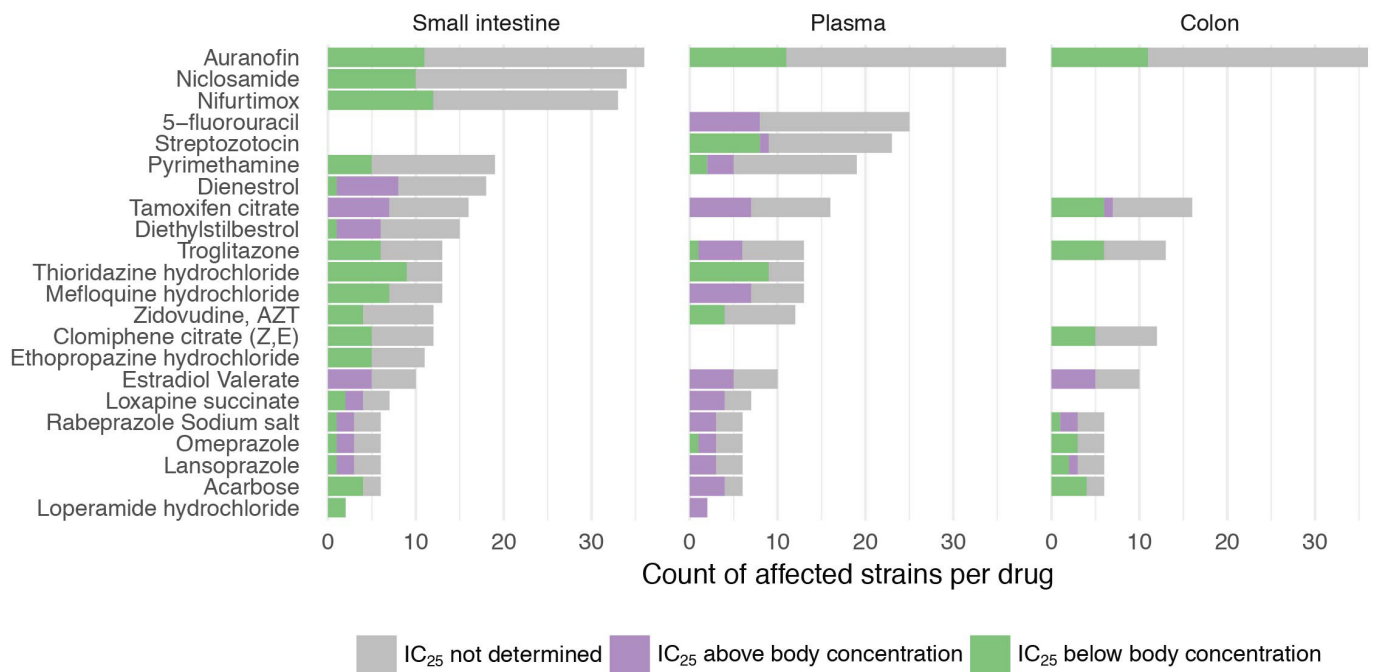
Extended Data Figure 4 | Effects of metformin in gut microbiota *in vivo* correlate with its *in vitro* activity. **a**, IC₂₅s of the antidiabetic drug metformin for a selection of 22 strains. Metformin did not inhibit any species in our screen as the concentration used, 20 μ M (red line), is below the IC₂₅ of all strains. However, at its estimated small intestinal and colon concentration of 1.5 mM (blue line), metformin would inhibit 3 of 22 tested strains. This exemplifies that more human-targeted drugs would

interfere with bacterial growth if doses were to be increased towards drug- and body-site-specific concentrations. **b**, IC₂₅s of metformin correlate well with its observed effects in humans⁶¹, based on the four species that overlapped between the two studies. Significant treatment effects on the species level were mapped to our set of strains for which we had determined IC₂₅s.



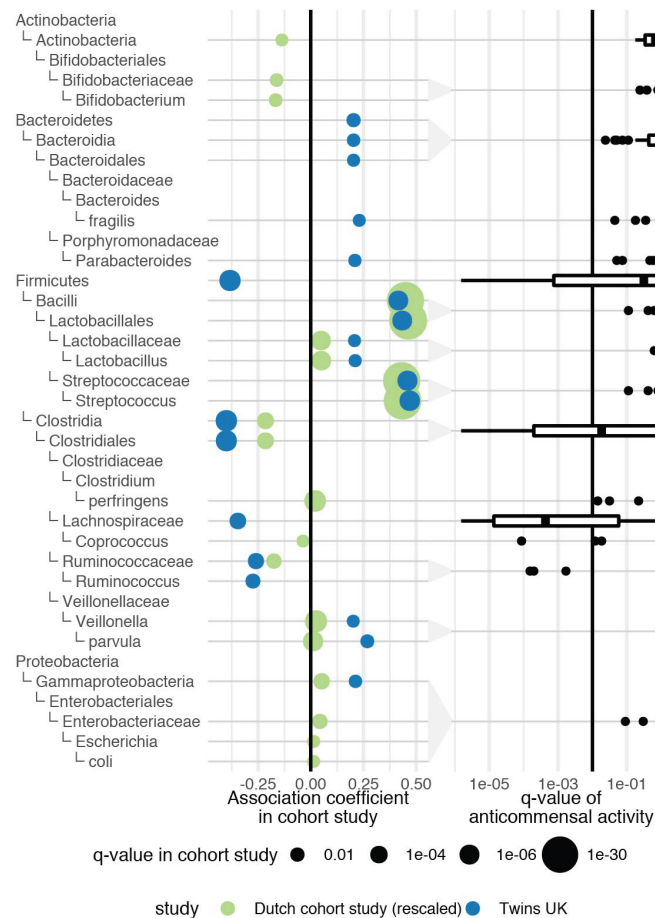
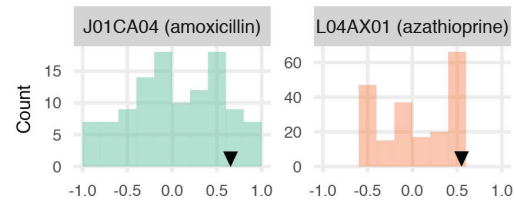
Extended Data Figure 5 | Validation of the screen and conservative hit-calling. **a, b**, Validation of our screen by IC₂₅ determination for 25 selected drugs in a subset of up to 27 strains reveals high precision (94%) and recall (85%). We considered IC₂₅ as the lowest concentration that reduces growth by at least 25% (see Methods). Breakdown into active and inactive compounds for drugs concentrations at the 20 μM concentration, used in our screen. True positives (TP), green; false positives (FP), red; true negatives (TN), grey; false negatives (FN), blue. **c**, Number of drugs with anticommensal activity versus the applied FDR threshold for all compounds (left) and human-targeted drugs (right). Increasing the FDR threshold from 0.01 to 0.1 (vertical grey lines) would nearly double the

fraction of drugs that affect human gut microbes. **d**, IC₂₅s of 25 drugs in up to 27 individual strains (see also **a, b**). The white area indicates the drug concentration range tested for each drug. Symbol sizes depict the number of strains with a particular IC₂₅, symbol colours indicate categorization into false negative, false positive, true negative and true positive, and symbol shapes qualify whether actual IC₂₅s were determined or IC₂₅ was deemed to be higher or lower than the highest or lowest concentration tested, respectively. Vertical line indicates the drug concentration used in screen (20 μM). IC₂₅s for all drug-strain pairs are listed in Supplementary Table 4. Particular drugs were responsible for false negatives in our screen (acarbose, loperamide, thioridazine), presumably owing to drug decay.

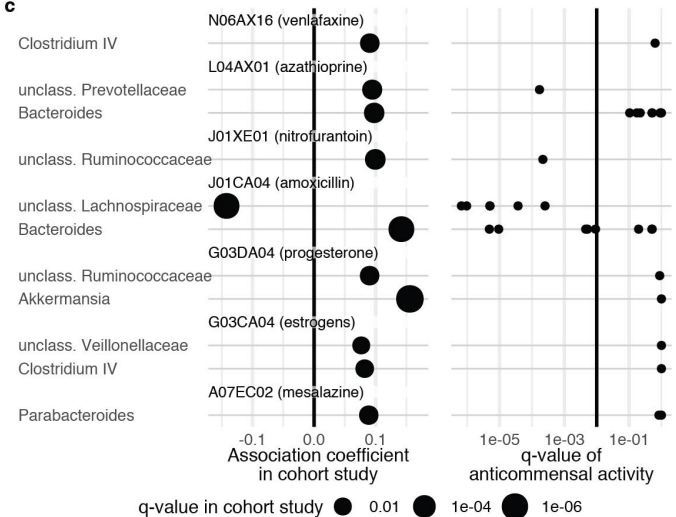
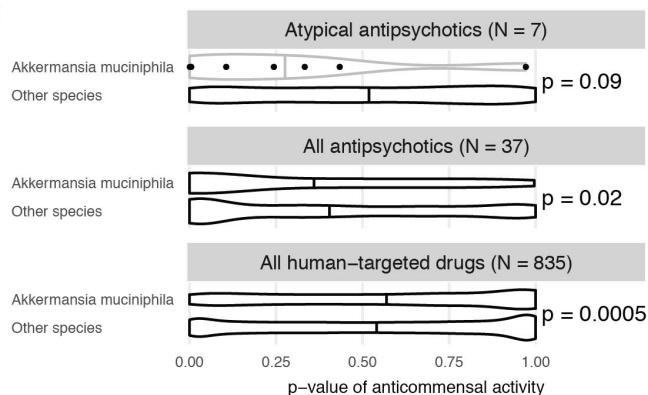


Extended Data Figure 6 | IC_{25} relation to drug concentrations in human body. For drug–strain pairs with measured IC_{25} s (see also Extended Data Fig. 5), we compared IC_{25} s with plasma and estimated small intestine and colon concentrations by plotting the number of strains that are affected in relation to whether they are above or below relevant body concentrations

(colour code). With the exception of oestradiol valerate and 5-FU (only plasma concentrations available), all other drugs with available body concentrations reach concentrations high enough in the body to reach their IC_{25} for at least one gut microbial species (out of up to 27 species tested for IC_{25} s).

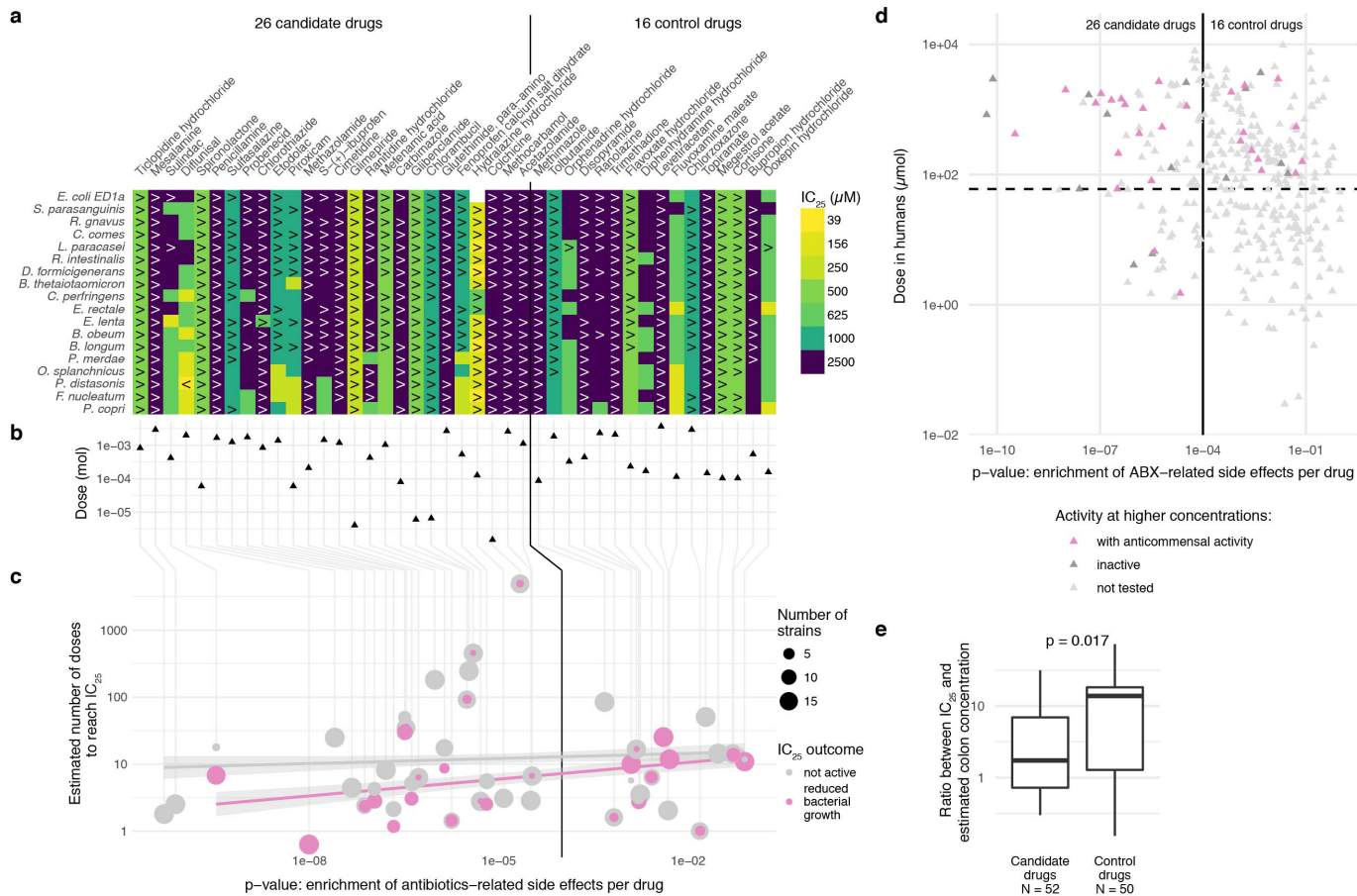
a Effect of proton-pump inhibitors *in vivo* and *in vitro***b** Flemish cohort studySpearman correlation between association coefficient and screen *p*-value

antibacterial drug with antimicrobial activity
human-targeted drug with antimicrobial activity

c**d****Extended Data Figure 7 | Concordance of drug *in vitro* species susceptibilities and drug-mediated shifts in microbiome composition of patients.**

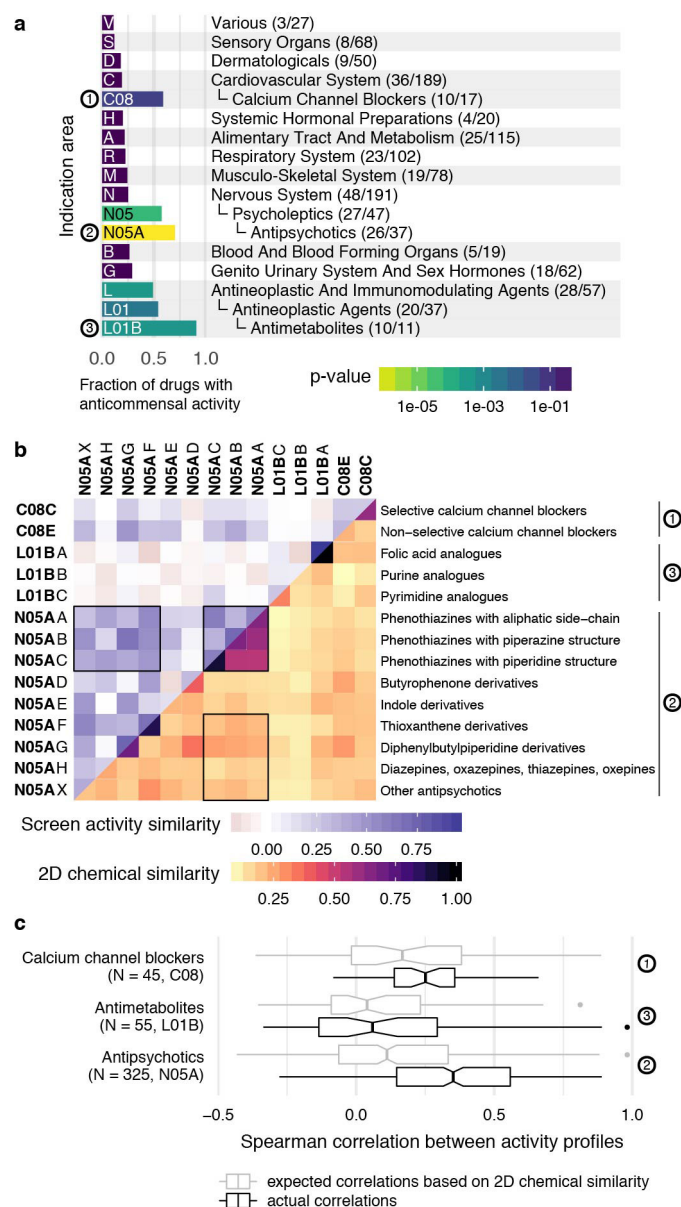
a, Association coefficients between PPI usage and relative taxonomic abundance in faecal microbiomes of PPI users from two studies (twins, UK cohort, green⁴; three independent cohorts from the Netherlands³, blue) (left) are compared to *in vitro* growth inhibition of isolates with same taxonomic rank in the presence of PPIs (omeprazole, lansoprazole and rabeprazole) as assessed by FDR-adjusted *P* values (*q* values) in our screen (right). Point size in the left panel corresponds to the *q* value as reported in the original study. Taxa that were reduced in patients (negative association coefficient, left of vertical black line) were mostly inhibited by PPIs in our screen (*q* value below 0.01, left of vertical black line), whereas enriched taxa were insensitive to PPIs. Box plots show: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 × IQR; points, outliers. For fewer than 10 data points, all points are shown individually. **b**, Spearman correlation coefficients between association coefficients of faecal microbiome composition after

consumption of amoxicillin or azathioprine⁷ and the screen *P* values. The histogram represents the background distribution of correlations between the *in vitro* data for all human-targeted drugs and the *in vivo* response to these drugs; correlations with amoxicillin or azathioprine are highlighted by triangles **c**, Comparisons between association coefficients and drugs from different therapeutic classes as assessed by Falony *et al.*⁷ and our *in vitro* data. **d**, A study of a cohort of patients with bipolar disease⁶ reported a significant decrease in abundance of *Akkermansia muciniphila* upon treatment with atypical antipsychotics (AAP). When we compared distributions of adjusted *P* values from our screen for different strains, *Akkermansia muciniphila* was significantly more sensitive than all other strains to antipsychotics in general and AAP in particular (*P* = 0.02 and *P* = 0.09, two-sided Wilcoxon rank sum test). By contrast, *A. muciniphila* is relatively more resistant than other strains across all human-targeted drugs (*P* = 0.0005, two-sided Wilcoxon rank sum test). Violin plot shows estimated density of points with the estimated median as vertical bar. For fewer than 10 data points, all points are also shown individually.



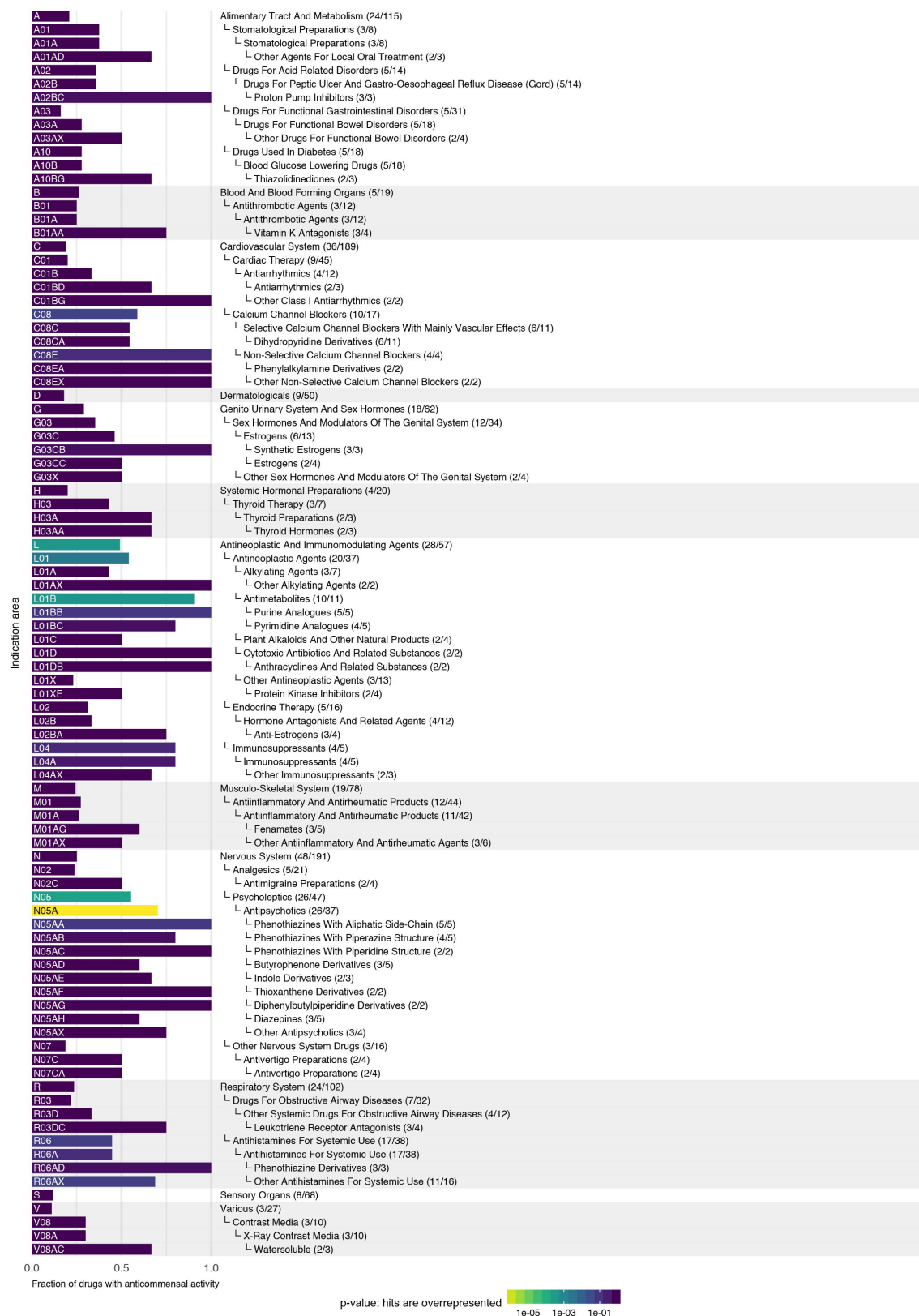
Extended Data Figure 8 | Evaluation of anticommensal activity predictions based on side-effects. **a**, IC₂₅ of 26 candidate compounds (*P* value for enrichment of antibiotic-related side effects $<1 \times 10^{-4}$, using a one-sided Fisher's exact test) and 16 control compounds (see also **d** of this Figure) were determined for 18 representative strains; results are depicted as an IC₂₅ heatmap. Drugs are ordered according to their similarity in side effects to antibiotics from left to right (for antibiotic-related side effects see Supplementary Table 5). Qualifiers indicate whether IC₂₅s are higher or lower than the indicated concentration; if no symbol, the box depicts the exact IC₂₅. If highest tested concentrations did not reduce growth of any of the tested strains, the compound was classified as inactive (for example, Topiramate). **b**, Dose of the tested compounds according to the Defined Daily Dose and Drugs@FDA databases (see also Supplementary Table 1). **c**, Based on a compound's recommended dose and its median IC₂₅ for different bacterial strains, we estimated the number of doses need to reach this IC₂₅. This number was plotted against the drug's *P* value for enrichment of antibiotic-related side effects. For direct comparison between the two groups, see Fig. 3c. Circles in magenta depict drug-strain pairs for which growth was reduced, showing a clear correlation between *P* values and the estimated number of doses (magenta line). To rule out the

possibility that the tested concentration range is causing this correlation, we also depict the estimated number of doses corresponding to the highest tested concentration (grey line), which exhibits no clear dependency between *P* value and number of doses. A vertical line across all panels connects all parameters attributable to a particular drug. **d**, Recommended single drug doses of human-targeted drugs with no anticommensal activity in our screen plotted against enrichment in antibiotic-related side effects (*n* = 339). Candidate and control drugs selection for testing for anticommensal activity at higher concentrations were selected on the basis of similarity to antibiotic-related side effects (vertical black line depicts prediction threshold) and aiming at drugs used at higher doses than concentration in our screen (horizontal dashed line). Purple and dark grey triangles indicate hits and non-hits from this validation effort, respectively. **e**, Ratios between IC₂₅ and estimated colon concentrations are significantly lower (*P* = 0.017, two-sided Wilcoxon rank sum test) for candidate drugs than for control drugs. For candidate drugs, 16 of 52 (31%) IC₂₅s were below the estimated colon concentrations while for control drugs this fraction was only 5 of 50 (10%). Box plots show: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 × IQR; points, outliers.



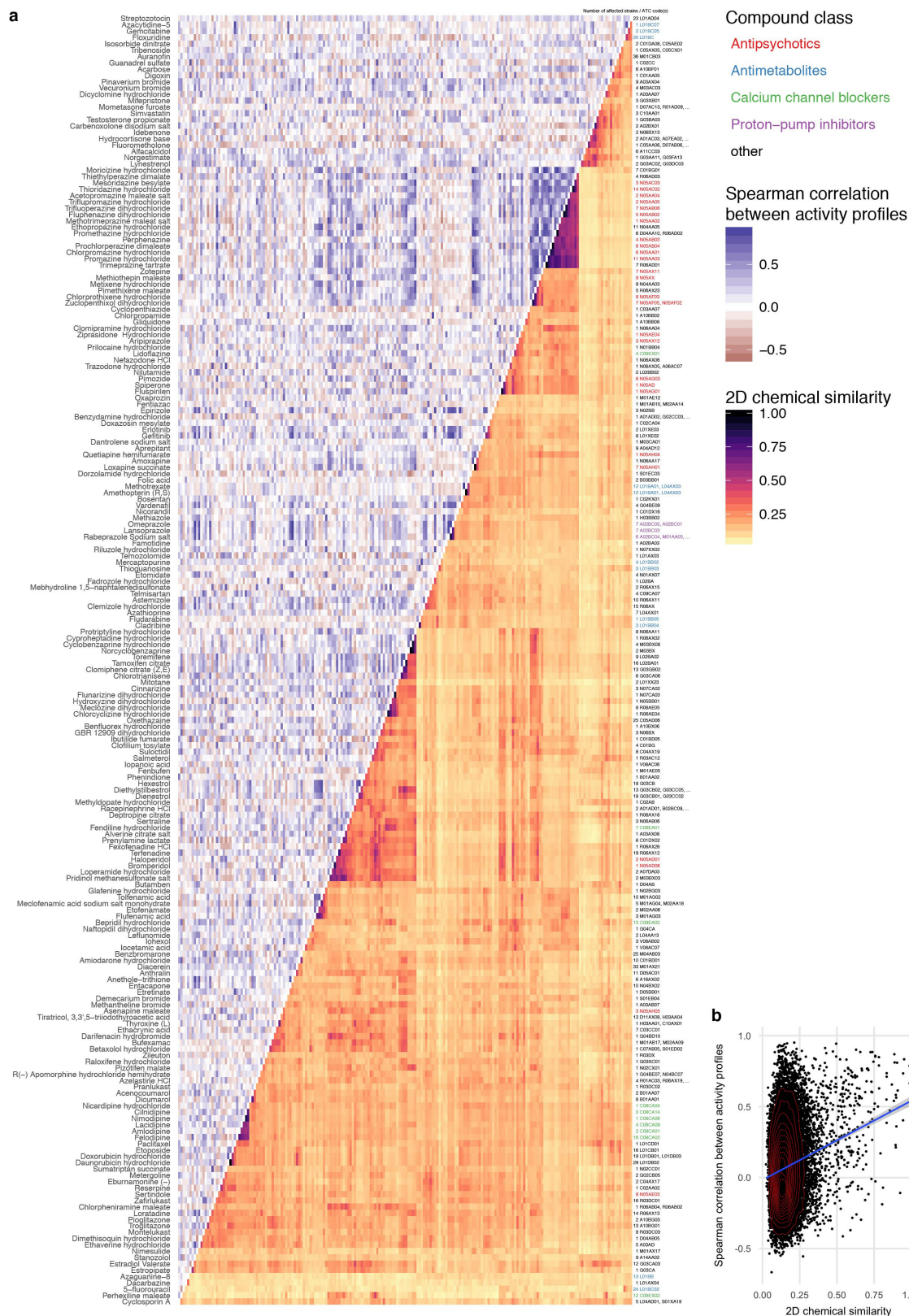
Extended Data Figure 9 | Drug therapeutic classes with anticommensal activity.

a, Fraction of drugs with anticommensal activity by ATC indication area (bars). All first-level indication areas and significantly enriched lower levels are shown (see also Extended Data Fig. 10). Significance (P value, one-sided Fischer's exact test) is controlled for multiple hypothesis testing (Benjamini–Hochberg) independently at each ATC hierarchy level. **b**, Heat map of anticommensal activity and chemical similarities of human-targeted drugs within the three significantly ATC indication levels from **a** (indicated by circled numbers). Colours represent the median of drug pairwise Spearman correlations within and between subgroups depicted, calculated from the growth profiles of the 40 strains in each drug (P values) or their chemical similarity (Tanimoto scores⁶²). Examples of structurally similar (phenothiazines; N05AA–AC) and diverse (N05AF–AX) antipsychotics that elicit similar responses in our screen are marked. **c**, Antipsychotics exhibit higher similarity in gut microbes they target than that expected on the basis of their structural similarity ($P = 2 \times 10^{-19}$ estimated from random permutations; other classes depicted show no significance difference). Box plots show: centre line, median; box limits, upper and lower quartiles; whiskers, $1.5 \times \text{IQR}$; points, outliers. Notches correspond roughly to a 95% confidence interval for comparing medians.



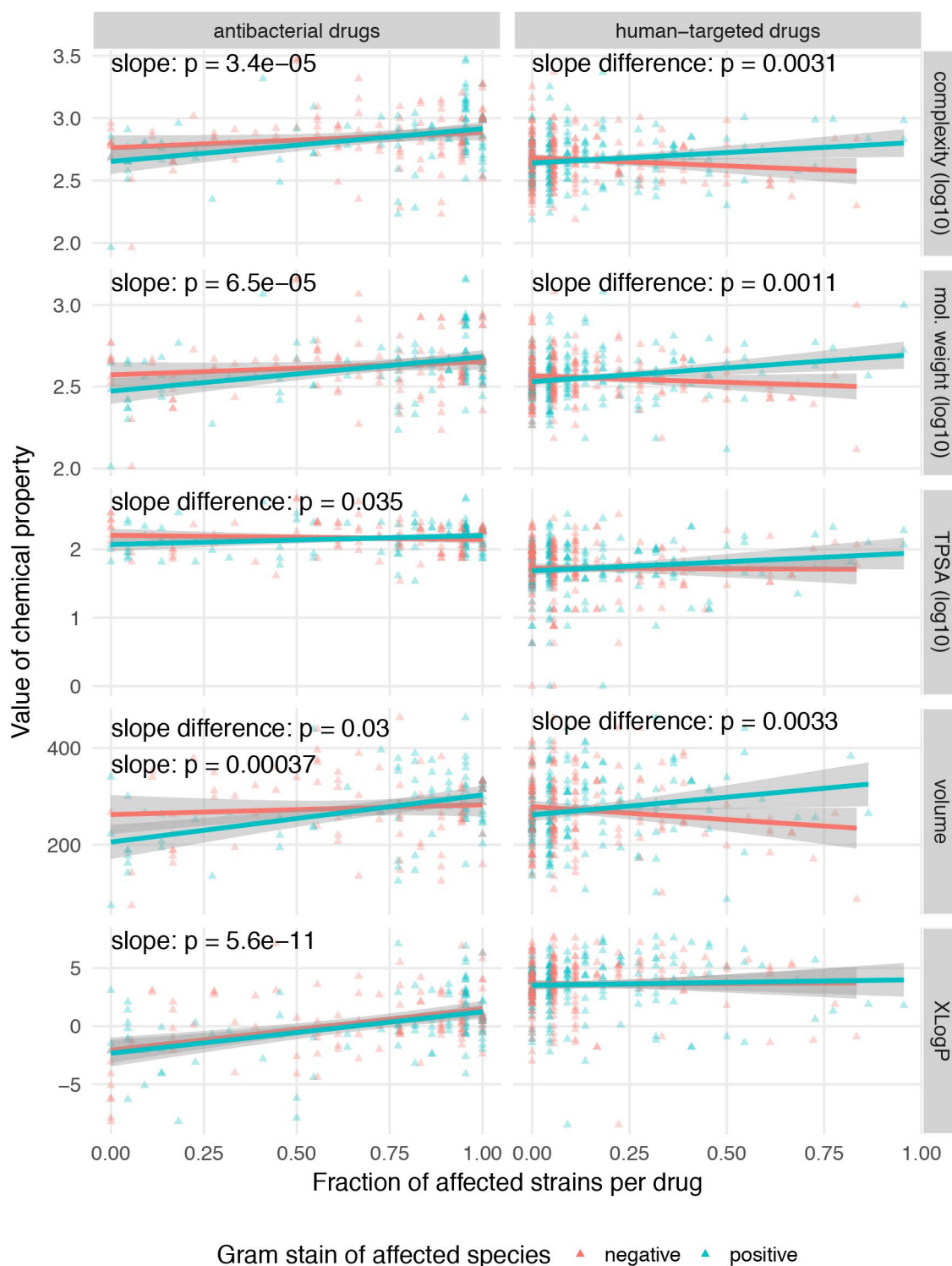
Extended Data Figure 10 | Drugs with anticomensal activity for all hierarchy levels of the ATC classification system. Fraction of drugs with anticomensal activity for all indication areas of the ATC classification scheme with a high fraction of active compounds. Shown are indication areas that contain at least two active compounds and a fraction of at least 50% active compounds, their parent terms and all top-level indication areas. Significance (*P* value, one-sided Fischer's exact test) is indicated by

the bar colour and corrected for multiple hypothesis testing (Benjamini-Hochberg) independently at each hierarchy level of the ATC. Many smaller classes, including PPIs (A02BC), non-selective calcium channel blockers (C08E), synthetic oestrogens (G03CB), leukotriene receptor antagonists (R03DC) and phenothiazine and other antihistamines (R06AD and R06AX) are enriched, but owing to multiple testing and the small numbers of drugs tested in each group, they do not reach a significant *P* value.



Extended Data Figure 11 | Comparing chemical similarity of drugs and similarity of hit profiles across gut microbes. **a**, Heat map of anticommensal activity and chemical similarities for all active human-targeted drugs in our screen. Drugs are clustered according to chemical similarity. Colours represent the median of drug pairwise Spearman correlations within and between subgroups depicted, calculated from the growth profiles of the 40 strains in each drug (P values) or their chemical similarity (Tanimoto scores⁶²). Several prominent groups are colour coded. Only drugs of some classes both share chemical similarity

and have similar effects on the 40 strains—for example, phenothiazine antipsychotics and antihistamines (N05A and R06AD), structurally similar dibenzothiazepines and dibenzoxazepines for antipsychotics and antidepressants (N05AH and N06AA), PPIs (A02BC), anti-oestrogens (L02BA), synthetic oestrogens (G03CB) and anti-inflammatory fenamates (M01AG and M02AA06). **b**, A mild correlation exists between chemical similarity (Tanimoto scores) and anticommensal activity similarity (drug pairwise Spearman correlations): $r_s = 0.12$ (P value of Spearman's test $< 2 \times 10^{-16}$).



Extended Data Figure 12 | More complex, bulkier and heavier human-targeted drugs are more effective against Gram-positive bacteria.

Fraction of inhibited Gram-positive (blue, $n = 22$) or Gram-negative (red, $n = 18$) strains per drug plotted against different chemical properties of the drugs. Chemical properties, such as complexity (based on atom types, symmetry, computed using the Bertz/Hendrickson/Ihlenfeldt formula), molecular weight, TPSA (an estimate of the area, in Å²), volume (in Å³) and XLogP (distribution coefficient that is a measure of differential

solubility in octanol and water) were obtained from PubChem⁶³. For each chemical property, we used a type II ANOVA to test for linear dependency between the fraction of affected species and the chemical property (slope). Additionally, we tested whether this dependency depended on the Gram stain (slope difference). It is possible that there is no significant slope without considering Gram stain, but that there is a significant difference between the slopes for the two Gram stains. Lines show a linear fit to the data, with 95% confidence intervals as shaded area.

A galaxy lacking dark matter

Pieter van Dokkum¹, Shany Danieli¹, Yotam Cohen¹, Allison Merritt^{1,2}, Aaron J. Romanowsky^{3,4}, Roberto Abraham⁵, Jean Brodie⁴, Charlie Conroy⁶, Deborah Lokhorst⁵, Lamiya Mowla¹, Ewan O'Sullivan⁶ & Jielai Zhang⁵

Studies of galaxy surveys in the context of the cold dark matter paradigm have shown that the mass of the dark matter halo and the total stellar mass are coupled through a function that varies smoothly with mass. Their average ratio $M_{\text{halo}}/M_{\text{stars}}$ has a minimum of about 30 for galaxies with stellar masses near that of the Milky Way (approximately 5×10^{10} solar masses) and increases both towards lower masses and towards higher masses^{1,2}. The scatter in this relation is not well known; it is generally thought to be less than a factor of two for massive galaxies but much larger for dwarf galaxies^{3,4}. Here we report the radial velocities of ten luminous globular-cluster-like objects in the ultra-diffuse galaxy⁵ NGC1052–DF2, which has a stellar mass of approximately 2×10^8 solar masses. We infer that its velocity dispersion is less than 10.5 kilometres per second with 90 per cent confidence, and we determine from this that its total mass within a radius of 7.6 kiloparsecs is less than 3.4×10^8 solar masses. This implies that the ratio $M_{\text{halo}}/M_{\text{stars}}$ is of order unity (and consistent with zero), a factor of at least 400 lower than expected². NGC1052–DF2 demonstrates that dark matter is not always coupled with baryonic matter on galactic scales.

NGC1052–DF2 was identified with the Dragonfly Telephoto Array⁶ in deep, wide-field imaging observations of the NGC 1052 group. The galaxy is not a new discovery; it was catalogued previously in a visual search of digitized photographic plates⁷. It stood out to us because of the remarkable contrast between its appearance in Dragonfly images and Sloan Digital Sky Survey (SDSS) data: with Dragonfly it is a low-surface-brightness object with some substructure and a spatial extent of about $2'$, whereas in SDSS it appears as a collection of point-like sources. Intrigued by the likelihood that these compact sources are associated with the low-surface-brightness object, we obtained follow-up spectroscopic observations of NGC1052–DF2 using the 10-m W. M. Keck Observatory. We also observed the galaxy with the Hubble Space Telescope (HST).

A colour image generated from the HST V_{606} and I_{814} data is shown in Fig. 1. The galaxy has a striking appearance. In terms of its apparent size and surface brightness, it resembles dwarf spheroidal galaxies such as those recently identified⁸ in the M101 group at 7 Mpc, but the fact that it is only marginally resolved implies that it is at a much greater distance. Using the I_{814} band image, we derived a surface-brightness-fluctuation distance of $D_{\text{SBF}} = 19.0 \pm 1.7$ Mpc (see Methods). It is located only $14'$ from the luminous elliptical galaxy NGC 1052, which has distance measurements ranging from 19.4 Mpc to 21.4 Mpc (refs 9, 10). We infer that NGC1052–DF2 is associated with NGC 1052, and we adopt $D \approx 20$ Mpc for the galaxy.

We parameterized the galaxy's structure with a two-dimensional Sérsic profile¹¹. The Sérsic index is $n = 0.6$, the axis ratio is $b/a = 0.85$, the central surface brightness is $\mu(V_{606}, 0) = 24.4$ mag arcsec^{−2}, and the effective radius along the major axis is $R_e = 22.6''$, or 2.2 kpc. We conclude that NGC1052–DF2 falls in the 'ultra-diffuse galaxy' (UDG) class⁵, which have $R_e > 1.5$ kpc and $\mu(g, 0) > 24$ mag arcsec^{−2}. In terms of its structural parameters, it is very similar to the galaxy Dragonfly 17

in the Coma cluster⁵. The total magnitude of NGC1052–DF2 is $M_{606} = -15.4$, and the total luminosity is $L_V = 1.1 \times 10^8$ solar luminosities, L_\odot . Its colour $V_{606} - I_{814} = 0.37 \pm 0.05$ in the AB magnitude system, similar to that of other UDGs and metal-poor globular clusters¹². The stellar mass was determined in two ways: by placing a stellar population at $D = 20$ Mpc that matches the global properties of NGC1052–DF2 (see Methods), and by assuming $M/L_V = 2.0$ as found for globular clusters¹³. Both methods give $M_{\text{stars}} \approx 2 \times 10^8$ solar masses, M_\odot .

We obtained spectroscopy of objects in the NGC1052–DF2 field with the W. M. Keck Observatory. Details of the observations and data reduction are given in the Methods section. We found ten objects with a radial velocity close to $1,800 \text{ km s}^{-1}$ (all other objects are Milky Way stars or background galaxies). We conclude that there is indeed a population of compact, luminous objects associated with NGC1052–DF2. Their spectra near the strongest calcium triplet lines are shown in Fig. 2. The mean velocity of the ten objects is $\langle v \rangle = 1,803^{+2}_{-2} \text{ km s}^{-1}$. The

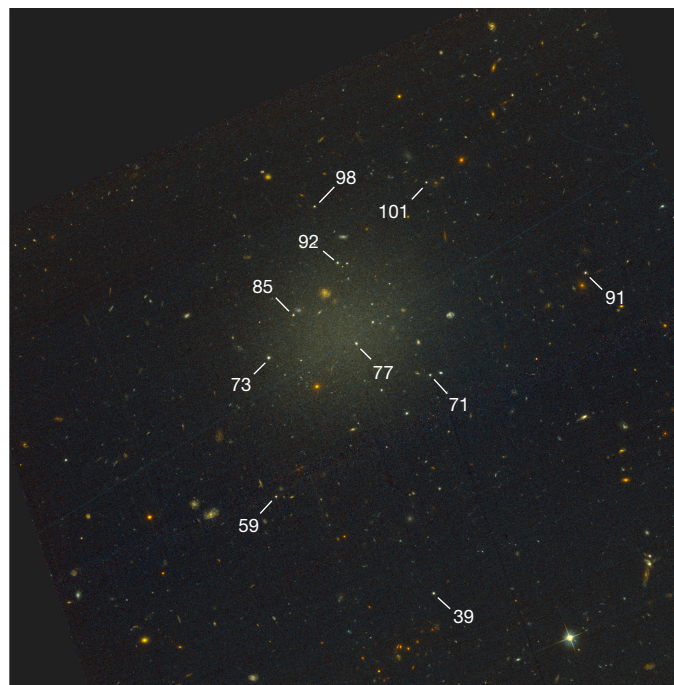


Figure 1 | HST/Advanced Camera for Surveys (ACS) image of NGC1052–DF2. NGC1052–DF2 was identified as a large (approximately $2'$) low-surface-brightness object, at right ascension $\alpha = 2 \text{ h } 41 \text{ min } 46.8 \text{ s}$, declination $\delta = -8^\circ 24' 12''$ (J2000). HST imaging of NGC1052–DF2 was obtained on 2016 November 10, using the ACS. The exposure time was 2,180 s in the V_{606} filter and 2,320 s in the I_{814} filter. The image spans $3.2' \times 3.2'$, or $18.6 \text{ kpc} \times 18.6 \text{ kpc}$ at the distance of NGC1052–DF2; north is up and east is to the left. Faint striping is caused by imperfect charge-transfer-efficiency removal. Ten spectroscopically confirmed luminous compact objects are marked.

¹Astronomy Department, Yale University, New Haven, Connecticut 06511, USA. ²Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany. ³Department of Physics and Astronomy, San Jose State University, San Jose, California 95192, USA. ⁴University of California Observatories, 1156 High Street, Santa Cruz, California 95064, USA. ⁵Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada. ⁶Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

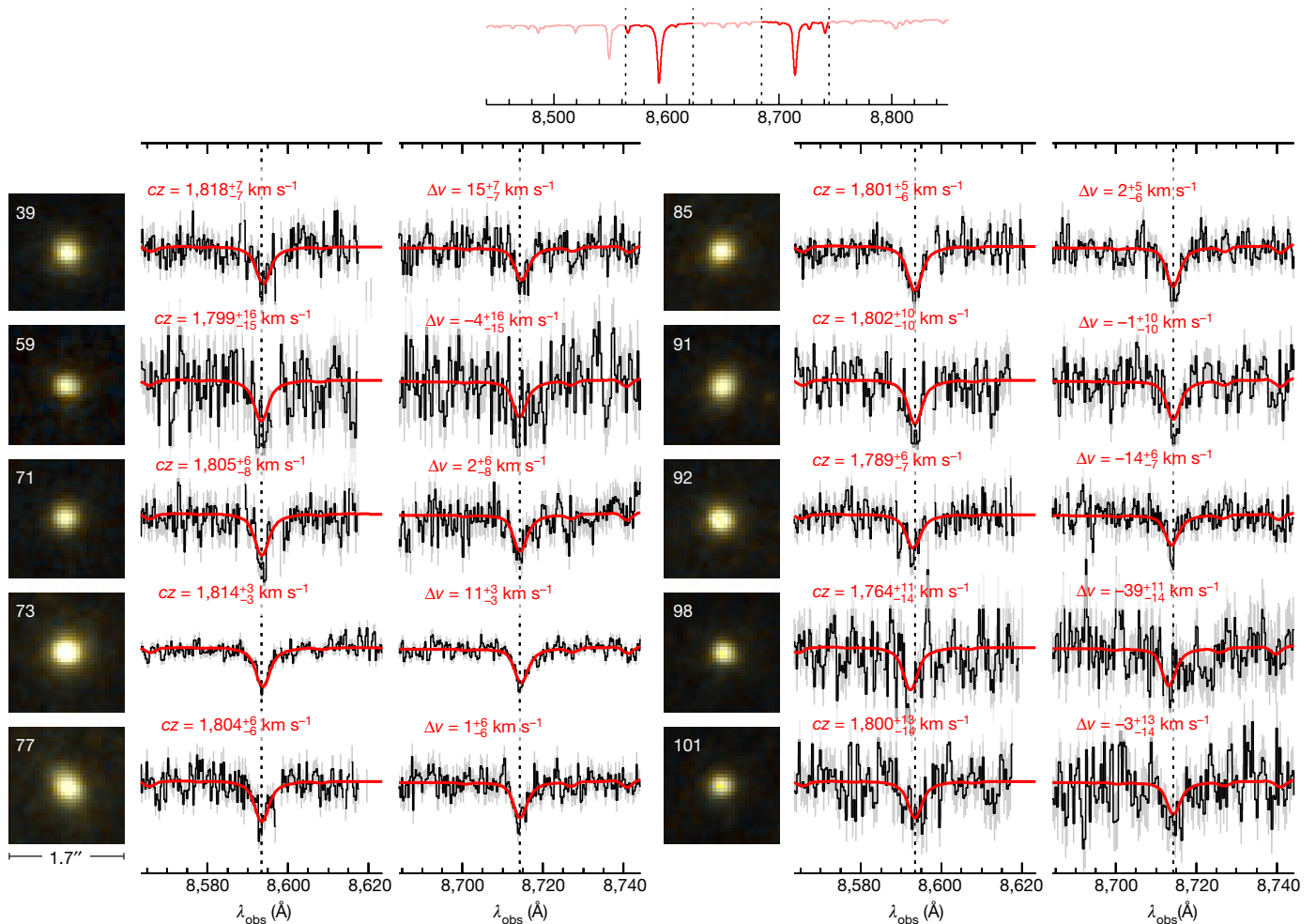


Figure 2 | Spectra of the compact objects. The square panels show the HST/ACS images of the ten confirmed compact objects. Each panel spans $1.7'' \times 1.7''$, or $165 \text{ pc} \times 165 \text{ pc}$ at the distance of NGC1052–DF2. The Keck spectra are shown next to the corresponding HST images. The regions around the reddest $\lambda 8,544.4$, $\lambda 8,664.5$ Ca II triplet lines are shown, as illustrated in the model spectrum at the top; the $\lambda 8,500.4$ Ca triplet line was included in the fit but falls on a sky line for the radial velocity of these objects. The spectra were obtained with the Low Resolution Imaging

Spectrometer (LRIS), the Deep Imaging Multi-Object Spectrograph (DEIMOS) or both. The spectral resolution is $\sigma_{\text{instr}} \approx 30 \text{ km s}^{-1}$. Uncertainties in the spectra are in grey. The signal-to-noise ratio ranges from 3.4 per pixel to 12.8 per pixel, with 0.4 \AA pixels. The red lines show the best-fitting models. Radial velocities cz are indicated, as well as the velocity offset with respect to the central $\langle v \rangle = 1,803 \text{ km s}^{-1}$. This velocity is indicated with dashed vertical lines.

NGC 1052 group has a radial velocity of $1,425 \text{ km s}^{-1}$, with a 1σ spread of only $\pm 111 \text{ km s}^{-1}$ (based on 21 galaxies). NGC1052–DF2 has a peculiar velocity of $+378 \text{ km s}^{-1}$ (3.4σ) with respect to the group and $+293 \text{ km s}^{-1}$ with respect to NGC 1052 itself (Fig. 3).

Images of the compact objects are shown in Fig. 2, and their locations are marked on Fig. 1. Their spatial distribution is somewhat more extended than that of the smooth galaxy light: their half-number radius is $R_{\text{gc}} \approx 3.1 \text{ kpc}$ (compared with effective radius $R_e = 2.2 \text{ kpc}$ for the light) and the outermost object is at $R_{\text{out}} = 7.6 \text{ kpc}$. In this respect, and in their compact morphologies (they are just-resolved in our HST images, as expected for their distance) and colours, they are similar to globular clusters and we will refer to them as such. Their luminosities are, however, much higher than those of typical globular clusters. The brightest (GC-73) has an absolute magnitude of $M_{606} = -10.1$, similar to that of the brightest globular cluster in the Milky Way (ω Centauri). Furthermore, the galaxy has no statistically significant population of globular clusters near the canonical peak of the luminosity function at $M_V \approx -7.5$. The properties of these enigmatic objects are the subject of another paper (P.v.D. *et al.*, manuscript in preparation).

The central observational result of the present study is the remarkably small spread among the velocities of the ten clusters (Fig. 3). The observed velocity dispersion is $\sigma_{\text{obs}} = 8.4 \text{ km s}^{-1}$, as measured with the

biweight estimator (see Methods). This value is much smaller than that in previously studied (cluster) UDGs^{12,14} and not much higher than the expectation from observational errors alone. Taking the errors into account, we find an intrinsic dispersion of $\sigma_{\text{intr}} = 3.2^{+5.5}_{-3.2} \text{ km s}^{-1}$. The 90% confidence upper limit is $\sigma_{\text{intr}} < 10.5 \text{ km s}^{-1}$. Within the Local Group, typical galaxies with velocity dispersions in this range are small ($R_e \approx 200 \text{ pc}$) and have a low stellar mass ($M_{\text{stars}} \approx (2-3) \times 10^6 M_{\odot}$)¹⁵. The average velocity dispersion of Local Group galaxies with $8.0 \leq \log(M/M_{\odot}) \leq 8.6$ is 32 km s^{-1} (dotted curve in Fig. 3a).

We calculate the corresponding 90% confidence upper limit on the mass of NGC1052–DF2 using the tracer mass estimator (TME) method¹⁶, which provides an estimate of the dynamical mass, M_{dyn} , within the radius of the outermost discrete tracer, R_{out} . We find $M_{\text{dyn}} < 3.4 \times 10^8 M_{\odot}$ within $R_{\text{out}} = 7.6 \text{ kpc}$. We also determine the dynamical mass within the projected half-number radius of the globular cluster system¹⁷ and find $M_{\text{dyn}} < 3.2 \times 10^8 M_{\odot}$ within $R_{\text{gc}} = 3.1 \text{ kpc}$.

In Fig. 4a, the enclosed mass is compared to the expected mass from the stars alone (orange line) and to models with different halo masses. The dynamical mass is consistent with the stellar mass, leaving little room for dark matter. The best fit to the kinematics is obtained for $M_{\text{halo}} = 0$, and the 90% confidence upper limit on the dark matter halo mass is $M_{\text{halo}} < 1.5 \times 10^8 M_{\odot}$. We note that the combination of the large

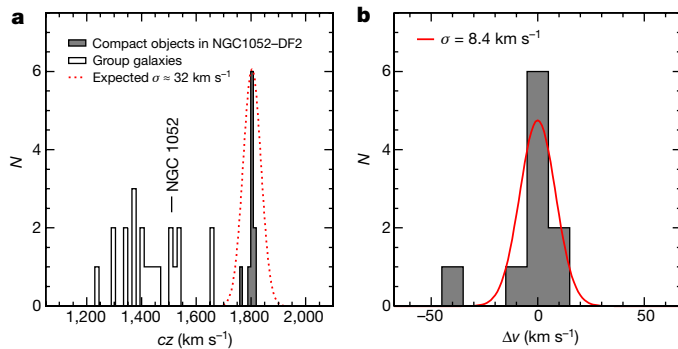


Figure 3 | Velocity dispersion. The filled grey histograms show the velocity distribution of the ten compact objects. **a**, Wide velocity range including the velocities of all 21 galaxies in the NASA/IPAC Extragalactic Database with $cz < 2,500 \text{ km s}^{-1}$ that are within a projected distance of two degrees from NGC 1052. The red dotted curve shows a Gaussian with a width of $\sigma = 32 \text{ km s}^{-1}$, the average velocity dispersion of Local Group galaxies with $8.0 \leq \log(M_{\text{stars}}/M_{\odot}) \leq 8.6$. **b**, Narrow velocity range centred on $cz = 1,803 \text{ km s}^{-1}$. The red solid curve is a Gaussian with a width that is equal to the biweight dispersion of the velocity distribution of the compact objects, $\sigma_{\text{obs}} = 8.4 \text{ km s}^{-1}$. Taking observational errors into account, we derive an intrinsic dispersion of $\sigma_{\text{intr}} = 3.2^{+5.5}_{-3.2} \text{ km s}^{-1}$, with the uncertainties 1 s.d. The 90% confidence upper limit on the intrinsic dispersion is $\sigma_{\text{intr}} < 10.5 \text{ km s}^{-1}$.

spatial extent and low dynamical mass of NGC1052–DF2 yields an unusually robust constraint on the total halo mass. Typically, kinematic tracers are only available out to a small fraction of the virial radius, and a large extrapolation is required to convert the measured enclosed mass to a total halo mass⁴. However, for a halo of mass $M_{200} \approx 10^8 M_{\odot}$, the virial radius is only about 10 kpc, similar to the radius at which the outermost globular clusters reside. As shown in Fig. 4b, a galaxy with a stellar mass of $M_{\text{stars}} = 2 \times 10^8 M_{\odot}$ is expected to have a halo mass of $M_{\text{halo}} \approx 6 \times 10^{10} M_{\odot}$, a factor of about 400 higher than the upper limit that we derive. We conclude that NGC1052–DF2 is extremely deficient in dark matter, and a good candidate for a ‘baryonic galaxy’ with no dark matter at all.

It is unknown how the galaxy was formed. One possibility is that it is an old tidal dwarf, formed from gas that was flung out of merging galaxies. Its location near an elliptical galaxy and its high peculiar velocity are consistent with this idea. Its relatively blue colour suggests a lower metallicity than might be expected for such objects¹⁸, but that depends on the detailed circumstances of its formation¹⁹. An alternative explanation is that the galaxy formed from low-metallicity gas that was swept up in quasar winds²⁰. The lack of dark matter, the location near a massive elliptical, the peculiar velocity and the colour are all qualitatively consistent with this scenario, although it is not clear whether the large size and low surface brightness of NGC1052–DF2 could have been produced by this process. A third option is that the galaxy formed from inflowing gas that fragmented before reaching NGC 1052, either relatively close to the assembling galaxy²¹ or out in the halo²². This fragmentation may have been aided or precipitated by jet-induced shocks²³. In any scenario, the luminous globular-cluster-like objects require an explanation; generically, it seems likely that the three peculiar aspects of the galaxy (its large size, its low dark matter content and its population of luminous compact objects) are related. An important missing piece of information is the number density of galaxies such as NGC1052–DF2. There are several other objects in our Cycle 24 HST programme that look broadly similar, but these do not have dynamical measurements yet—and the fact that other UDGs have anomalously high, rather than low, dark matter fractions^{12,14} demonstrates that such data are needed to interpret these galaxies.

Regardless of the formation history of NGC1052–DF2, its existence has implications for the dark matter paradigm. Our results demonstrate that dark matter is separable from galaxies, which is (under certain

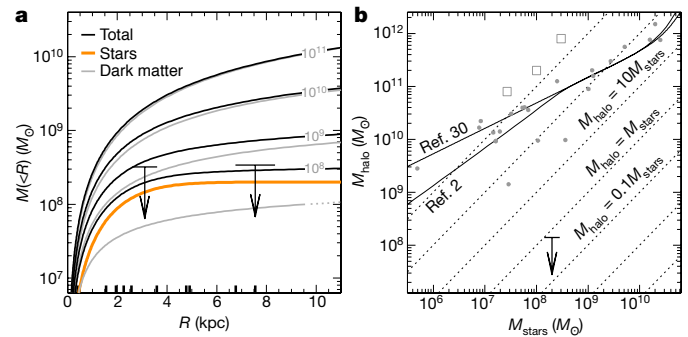


Figure 4 | Constraints on the halo mass. **a**, Enclosed mass profiles for Navarro–Frenk–White haloes²⁹ of masses $M_{200} = 10^8 M_{\odot}$, $10^9 M_{\odot}$, $10^{10} M_{\odot}$ and $10^{11} M_{\odot}$ (grey lines). The $10^8 M_{\odot}$ halo profile is shown by a dotted line beyond $R = R_{200} \approx 10 \text{ kpc}$. The orange curve is the enclosed mass profile for the stellar component, and the black curves are the total mass profiles $M_{\text{total}} = M_{\text{stars}} + M_{\text{halo}}$. The ten globular clusters are at distances ranging from $R = 0.4 \text{ kpc}$ to $R = 7.6 \text{ kpc}$; short vertical bars on the horizontal axis indicate the locations of individual clusters. The 90% upper limits on the total enclosed mass of NGC1052–DF2 are shown by arrows. The limit at $R = 7.6 \text{ kpc}$ was determined with the TME method¹⁶. The arrow at $R = 3.1 \text{ kpc}$ is the mass limit within the half-number radius of the globular cluster system¹⁷. The dynamical mass of NGC1052–DF2 is consistent with the stellar mass and leaves little room for a dark matter halo. **b**, The upper limit on the halo mass and comparison to the expected dark matter mass from studies that model the halo mass function and the evolution of galaxies^{2,30}. Grey solid symbols are nearby dwarf galaxies with rotation curves extending to at least two disk scale lengths⁴. Open squares are three cluster UDGs with measured kinematics: VCC 1287¹⁴, Dragonfly 44¹² and DFX1¹². NGC1052–DF2 falls a factor of at least 400 below the canonical relations.

circumstances) expected if it is bound to baryons through nothing but gravity. The ‘bullet cluster’ demonstrates that dark matter does not always trace the bulk of the baryonic mass²⁴, which in clusters is in the form of gas. NGC1052–DF2 enables us to make the complementary point that dark matter does not always coincide with galaxies either: it is a distinct ‘substance’ that may or may not be present in a galaxy. Furthermore, and paradoxically, the existence of NGC1052–DF2 may falsify alternatives to dark matter. In theories such as modified Newtonian dynamics (MOND)²⁵ and the recently proposed emergent gravity paradigm²⁶, a ‘dark matter’ signature should always be detected, as it is an unavoidable consequence of the presence of ordinary matter. In fact, it had been argued previously²⁷ that the apparent absence of galaxies such as NGC1052–DF2 constituted a falsification of the standard cosmological model and offered evidence for modified gravity. For a MOND acceleration scale of $a_0 = 3.7 \times 10^3 \text{ km}^2 \text{ s}^{-2} \text{ kpc}^{-1}$, the expected²⁸ velocity dispersion of NGC1052–DF2 is $\sigma_M \approx (0.05 GM_{\text{stars}} a_0)^{1/4} \approx 20 \text{ km s}^{-1}$, where G is the gravitational constant—a factor of two higher than the 90% upper limit on the observed dispersion.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 October 2017; accepted 8 January 2018.

1. Moster, B. P. et al. Constraints on the relationship between stellar mass and halo mass at low and high redshift. *Astrophys. J.* **710**, 903–923 (2010).
2. Behroozi, P. S., Wechsler, R. H. & Conroy, C. The average star formation histories of galaxies in dark matter halos from $z = 0$ –8. *Astrophys. J.* **770**, 57 (2013).
3. More, S. et al. Satellite kinematics—III. Halo masses of central galaxies in SDSS. *Mon. Not. R. Astron. Soc.* **410**, 210–226 (2011).
4. Oman, K. A. et al. Missing dark matter in dwarf galaxies? *Mon. Not. R. Astron. Soc.* **460**, 3610–3623 (2016).
5. van Dokkum, P. G. et al. Forty-seven Milky Way-sized, extremely diffuse galaxies in the Coma cluster. *Astrophys. J.* **798**, L45 (2015).
6. Abraham, R. G. & van Dokkum, P. G. Ultra-low surface brightness imaging with the Dragonfly telephoto array. *Publ. Astron. Soc. Pac.* **126**, 55–69 (2014).
7. Karachentsev, I. D., Karachentseva, V. E., Suchkov, A. A. & Grebel, E. K. Dwarf galaxy candidates found on the SERC EJ sky survey. *Astron. Astrophys. Suppl. Ser.* **145**, 415–423 (2000).

8. Danieli, S. *et al.* The Dragonfly Nearby Galaxies Survey. III. The luminosity function of the M101 group. *Astrophys. J.* **837**, 136 (2017).
9. Tonry, J. L. *et al.* The SBF survey of galaxy distances. IV. SBF magnitudes, colors, and distances. *Astrophys. J.* **546**, 681–693 (2001).
10. Blakeslee, J. P., Lucey, J. R., Barris, B. J., Hudson, M. J. & Tonry, J. L. A synthesis of data from fundamental plane and surface brightness fluctuation surveys. *Mon. Not. R. Astron. Soc.* **327**, 1004–1020 (2001).
11. Peng, C. Y., Ho, L. C., Impey, C. D. & Rix, H.-W. Detailed structural decomposition of galaxy images. *Astron. J.* **124**, 266–293 (2002).
12. van Dokkum, P. *et al.* Extensive globular cluster systems associated with ultra diffuse galaxies in the Coma cluster. *Astrophys. J.* **844**, L11 (2017).
13. McLaughlin, D. E. & van der Marel, R. P. Resolved massive star clusters in the Milky Way and its satellites: brightness profiles and a catalog of fundamental parameters. *Astrophys. J. Suppl. Ser.* **161**, 304–360 (2005).
14. Beasley, M. A. *et al.* An overmassive dark halo around an ultra-diffuse galaxy in the Virgo cluster. *Astrophys. J.* **819**, L20 (2016).
15. McConnachie, A. W. The observed properties of dwarf galaxies in and around the Local Group. *Astron. J.* **144**, 4 (2012).
16. Watkins, L. L., Evans, N. W. & An, J. H. The masses of the Milky Way and Andromeda galaxies. *Mon. Not. R. Astron. Soc.* **406**, 264–278 (2010).
17. Wolf, J. *et al.* Accurate masses for dispersion-supported galaxies. *Mon. Not. R. Astron. Soc.* **406**, 1220–1237 (2010).
18. Bournaud, F. *et al.* Missing mass in collisional debris from galaxies. *Science* **316**, 1166 (2007).
19. Ploekinger, S. *et al.* Tidal dwarf galaxies in cosmological simulations. *Mon. Not. R. Astron. Soc.* **474**, 580–596 (2018).
20. Natarajan, P., Sigurdsson, S. & Silk, J. Quasar outflows and the formation of dwarf galaxies. *Mon. Not. R. Astron. Soc.* **298**, 577–582 (1998).
21. Canning, R. E. A. *et al.* Filamentary star formation in NGC 1275. *Mon. Not. R. Astron. Soc.* **444**, 336–349 (2014).
22. Mandelker, N., van Dokkum, P. G., Brodie, J. P. & Ceverino, D. Cold filamentary accretion and the formation of metal poor globular clusters. Preprint at <https://arxiv.org/abs/1711.09108> (2017).
23. van Breugel, W., Filippenko, A. V., Heckman, T. & Miley, G. Minkowski's object—a starburst triggered by a radio jet. *Astrophys. J.* **293**, 83–93 (1985).
24. Clowe, D. *et al.* A direct empirical proof of the existence of dark matter. *Astrophys. J.* **648**, L109–L113 (2006).
25. Milgrom, M. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astrophys. J.* **270**, 365–370 (1983).
26. Verlinde, E. P. Emergent gravity and the dark universe. *SciPost Phys.* **2**, 016 (2017).
27. Kroupa, P. The dark matter crisis: falsification of the current standard model of cosmology. *Publ. Astron. Soc. Aust.* **29**, 395–433 (2012).
28. Angus, G. W. Dwarf spheroidals in MOND. *Mon. Not. R. Astron. Soc.* **387**, 1481–1488 (2008).
29. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
30. Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V. & Faber, S. M. Constraining the galaxy–halo connection over the last 13.3 Gyr: star formation histories, galaxy mergers and structural properties. *Mon. Not. R. Astron. Soc.* **470**, 651–687 (2017).

Acknowledgements A.J.R. was supported by National Science Foundation grant AST-1616710 and as a Research Corporation for Science Advancement Cottrell Scholar. Results are based on observations obtained with the W. M. Keck Observatory on Mauna Kea, Hawaii. We are grateful for the opportunity to conduct observations from this mountain and wish to acknowledge its important cultural role within the indigenous Hawaiian community.

Author Contributions P.v.D. led the observations, data reduction and analysis, and wrote the manuscript. S.D. visually identified the galaxy in the Dragonfly data and created the model galaxies to determine the distance. Y.C. measured the structural parameters of the object. A.M. used an automated approach to verify the visual detections of low-surface-brightness galaxies in the Dragonfly data. J.Z. and A.M. reduced the Dragonfly data. E.O'S. provided the MMT image. All authors contributed to aspects of the analysis and to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to P.v.D. (pieter.vandokkum@yale.edu).

METHODS

Imaging. In this paper, we use imaging from the Dragonfly Telephoto Array, the Sloan Digital Sky Survey, the MMT, the Gemini North telescope and the Hubble Space Telescope.

Dragonfly. The Dragonfly Telephoto Array⁶ data were obtained in the context of the Dragonfly Nearby Galaxy Survey³¹. Dragonfly was operating with eight telephoto lenses at the time of the observations, forming the optical equivalent of an $f/1.0$ refractor with a 40-cm aperture. The data reach a 1σ surface brightness limit of $\mu(g) \approx 29$ mag arcsec⁻² in $12'' \times 12''$ boxes³¹. The full Dragonfly field is shown in Extended Data Fig. 1, as well as the area around NGC 1052 and NGC1052–DF2. SDSS. SDSS images in the g , r and i bands were obtained from the DR14 Sky Server³². To generate the image in Extended Data Fig. 4, the data in the three bands were summed without weighting or scaling. The object is located near the corner of a frame. We note that the SDSS photometry for the compact objects is not reliable, as it is ‘contaminated’ by the low surface brightness emission of the galaxy (which is just detected in SDSS).

MMT. MMT/Megacam imaging of the NGC 1052 field was available from a project to image the globular cluster systems of nearby early-type galaxies³³. The data were taken in the r and i bands, in $0.9''$ seeing. They were used for target selection in our first Keck spectroscopic run.

Gemini. We obtained imaging with the Gemini-North Multi Object Spectrograph³⁴ (GMOS) in programme GN-2016B-DD-3. The observations were made on 2016 October 10, with total exposure times of 3,000 s in the g band and 3,000 s in the i band. The seeing was $0.65''$ in i and $0.70''$ in g . The data were reduced using the Gemini IRAF package. Low-order polynomials were fitted to individual (dithered) 300-s frames after carefully masking objects, to reduce large-scale background gradients at low surface brightness levels. The images were used to aid in the target selection and mask design for the LRIS spectroscopy. The Gemini data also provide the best available information on the regularity of the galaxy at low surface brightness levels (see below). The combined frames still show some background artefacts, but they are less prominent than those in the HST data. Finally, a visual inspection of the Gemini images prompted us to request a change in the scheduling of HST program GO-14644, moving the already-planned ACS observation of NGC1052–DF2 to an earlier date.

HST. The HST data were obtained as part of programme GO-14644. The aim of this programme is to obtain ACS images of a sample of 23 low-surface-brightness objects that were identified in fields of the Dragonfly Nearby Galaxy Survey³¹. NGC1052–DF2 was observed on 2016 November 16, for a total of two orbits. Exposure times were 2,180 s in V_{606} and 2,320 s in I_{814} . In this paper we use images processed by the drizzle (variable-pixel linear reconstruction) method, which have been corrected for charge-transfer efficiency (CTE) effects. Despite this correction, some CTE artefacts are still visible in the data (see Fig. 1).

Structural parameters. The size, surface brightness and other structural parameters of NGC1052–DF2 were determined from the HST data. First, the I_{814} image was rebinned to a lower resolution to increase the signal-to-noise (S/N) ratio per pixel. A preliminary object mask was then created from a segmentation map produced by SExtractor³⁵, using a relatively high detection threshold. A first-pass Sérsic model³⁶ for the galaxy was obtained using the GALFIT software¹¹. This model was subtracted from the data, and an improved object mask was created using a lower SExtractor detection threshold. Finally, GALFIT was run again to obtain the final structural parameters and total magnitude. The total magnitude in the V_{606} band, and the $V_{606} - I_{814}$ colour, were determined by running GALFIT on the (binned and masked) V_{606} image with all parameters except the total magnitude fixed to the I_{814} values. The structural parameters, total magnitude and colour are listed in the main text. We note that we measured nearly identical structural parameters from the Gemini images.

Spectroscopy. We obtained spectroscopy of compact objects in the NGC1052–DF2 field in two observing runs. The first set of observations was obtained on 2016 September 28 with DEIMOS³⁷ on Keck II, and the second was obtained on 2016 October 26 and 27 with LRIS³⁸ on Keck I.

DEIMOS observations. Conditions were variable, with cirrus clouds increasing throughout the night. We obtained 4 h of total on-source exposure time on a single multi-object mask; a second mask was exposed but yielded no useful data owing to clouds. The target selection algorithm gave priority to compact objects with $i < 22.5$ near NGC1052–DF2, selected from the MMT data. We used the 1,200 lines mm⁻¹ grating with a slit width of $0.75''$, providing an instrumental resolution of $\sigma_{\text{instr}} \approx 25$ km s⁻¹. The data were reduced with the same pipeline that we used previously^{12,39} for the Coma UDGs Dragonfly 44 and DFX1. The globular clusters GC-39, GC-71, GC-73, GC-77, GC-85, GC-92 and GC-98 (see Figs 1 and 2) were included in this mask.

LRIS observations. Two multi-slit masks were observed, one for 3.5 h (on source) on October 26 and a second for 4 h on October 27. Targets were selected from the Gemini data, giving priority to compact objects that had not been observed

with DEIMOS. Conditions were fair during both nights, with intermittent cirrus and seeing of about $1''$. We used the gold-coated grating with 1,200 lines mm⁻¹ blazed at $9,000 \text{ \AA}$. The instrumental resolution σ_{instr} is about 30 km s^{-1} . A custom pipeline was used for the data reduction, modelled on the one that we developed for DEIMOS. LRIS suffers from considerable flexure, and the main difference from the DEIMOS pipeline is that each individual 1,800-s exposure was reduced and calibrated independently to avoid smoothing of the combined spectra in the wavelength direction. The clusters GC-39, GC-59, GC-73, GC-91 and GC-101 were included in the first mask; GC-39, GC-71, GC-77, GC-85 and GC-92 were included in the second.

Combined spectra. Most compact objects were observed multiple times, and we combined these individual spectra to increase the S/N ratio. All spectra were given equal weight, and before combining they were divided by a low-order polynomial fit to the continuum in the calcium triplet (CaT) region. We tested that weighting by the formal S/N ratio instead does not change the best-fit velocities. The individual spectra were also shifted in wavelength to account for the heliocentric correction; this needs to be done at this stage as the DEIMOS and LRIS data were taken one month apart. Six GCs have at least two independent observations and effective exposure times of about 8 h; four were observed only once: GC-59, GC-91, GC-98 and GC-101. The S/N ratio in the final spectra ranges from 3–4 per pixel for GC-59, GC-98 and GC-101 to 13 per pixel for GC-73. A pixel is 0.4 \AA or 14 km s^{-1} .

Velocity measurements. Radial velocities were determined for all objects with detected CaT absorption lines. No fits were attempted for background galaxies (based on the detection of redshifted emission lines), Milky Way stars or spectra with no visible features. The measurements were performed by fitting a template spectrum to the observations, using the emcee MCMC algorithm⁴⁰. The template is a high-resolution model of a stellar population⁴¹, smoothed to the instrumental resolution. The model has an age of 11 Gyr and a metallicity of $[\text{Fe}/\text{H}] = -1$, which is consistent with the colours of the compact objects ($V_{606} - I_{814} \approx 0.35$); the results are independent of the precise choice of template. The fits are performed over the observed wavelength range $8,530 \text{ \AA} \leq \lambda \leq 8,750 \text{ \AA}$ and have three free parameters: the radial velocity, the normalization and an additive term that serves as a template mismatch parameter (as it allows the strength of the absorption lines to vary with respect to the continuum). The fit was performed twice. After the first pass, all pixels that deviate by more than 3σ were masked in the second fit. This step reduces the effect of systematic sky-subtraction residuals on the fit.

The uncertainties given by the emcee method do not take systematic errors into account. Following previous studies⁴², we determined the uncertainties in the velocities by shuffling the residuals. For each spectrum, the best-fitting model was subtracted from the data. Next, 500 realizations of the data were created by randomizing the wavelengths of the residual spectra and then adding the shuffled residuals to the best-fitting model. These 500 spectra were then fitted by using a simple χ^2 minimization, and the 16th and 84th percentiles of the resulting velocity distribution yield the error bars. To preserve the higher noise at the location of sky lines, the randomization was done separately for pixels at the locations of sky lines and for pixels in between the lines. We find that the resulting errors show the expected inverse trend with the S/N ratio of the spectra, whereas the emcee errors show large variation at fixed S/N ratio.

We tested the reliability of the errors by applying the same procedure to the individual LRIS and DEIMOS spectra for the six objects that were observed with both instruments (GC-39, GC-71, GC-73, GC-77, GC-85 and GC-92). For each object, the observed difference between the LRIS and DEIMOS velocities was divided by the expected error in the difference. The root mean square (r.m.s.) of these ratios is 1.2 ± 0.3 , that is, the empirically determined uncertainties are consistent with the observed differences between the independently measured LRIS and DEIMOS velocities.

Velocity dispersion. The observed velocity distribution of the ten clusters is not well approximated by a Gaussian. Six of the ten have velocities that are within $\pm 4 \text{ km s}^{-1}$ of the mean and one is 39 km s^{-1} removed from the mean. As a result, different ways to estimate the Gaussian-equivalent velocity spread σ_{obs} yield different answers. The normalized median absolute deviation $\sigma_{\text{obs,nmad}} = 4.7 \text{ km s}^{-1}$, the biweight⁴³ $\sigma_{\text{obs,bi}} = 8.4 \text{ km s}^{-1}$ and the r.m.s. $\sigma_{\text{obs,rms}} = 14.3 \text{ km s}^{-1}$. The r.m.s. is driven by one object with a relatively large velocity uncertainty (GC-98) and is inconsistent with the velocity distribution of the other nine. Specifically, for ten objects drawn from a Gaussian distribution and including the observed errors, the probability of measuring $\sigma_{\text{bi}} \leq 8.4$ if $\sigma_{\text{rms}} \geq 14.3$ is 1.5%, and the probability of measuring $\sigma_{\text{nmad}} \leq 4.7$ is 3×10^{-3} . We therefore adopt the biweight dispersion rather than the r.m.s. when determining the intrinsic dispersion below. We then show that the presence of GC-98 is consistent with the intrinsic dispersion that we derive using this statistic.

The observed dispersion must be corrected for observational errors, which are of the same order as σ_{obs} itself. We determined the intrinsic dispersion and its uncertainty in the following way. For a given value of σ_{test} , we generated

1,000 samples of ten velocities, distributed according to a Gaussian of width σ_{test} . The ten velocities in each sample were then perturbed with errors, drawn from Gaussians with widths equal to the empirically determined uncertainties in the measured dispersions. Using the biweight estimator, ‘measured’ dispersions $\sigma_{\text{obs,test}}$ were calculated for all samples. If the value 8.4 is within the 16th to 84th percentiles of the distribution of $\sigma_{\text{obs,test}}$ then σ_{test} is within the $\pm 1\sigma$ uncertainty on σ_{intr} . This method gives $\sigma_{\text{intr}} = 3.2^{+5.5}_{-3.2} \text{ km s}^{-1}$. As the intrinsic dispersion is consistent with zero, a more meaningful number than the best-fit is the 90% confidence upper limit; we find $\sigma_{\text{intr}} < 10.5 \text{ km s}^{-1}$.

We now return to the question whether GC-98 is consistent with the other nine objects. This cluster has a velocity offset of $\Delta v = -39^{+11}_{-14} \text{ km s}^{-1}$. For the upper limit on the intrinsic dispersion ($\sigma_{\text{intr}} = 10.5 \text{ km s}^{-1}$) the object is a 2.4σ outlier, and the probability of having a more than 2.4σ outlier in a sample of ten is 15%. Interestingly, the combination of the biweight constraint of $\sigma_{\text{intr}} < 10.5$ and the existence of GC-98 implies a fairly narrow range of intrinsic dispersions that are consistent with the entire set of ten velocities (assuming that they are drawn from a Gaussian distribution and that the errors are correct). The probability of having at least one object with the velocity of GC-98 is $< 10\%$ if the intrinsic dispersion is $\sigma_{\text{intr}} < 8.8 \text{ km s}^{-1}$. Taking both 90% confidence limits at face value, the allowed range in the intrinsic dispersion is $8.8 \text{ km s}^{-1} < \sigma_{\text{intr}} < 10.5 \text{ km s}^{-1}$.

Expected dispersion from Local Group galaxies. In Fig. 3a we illustrate how unusual the kinematics of NGC1052–DF2 are by comparing the observed velocity distribution to that expected from Local Group galaxies with the same stellar mass (broken red curve). The width of this Gaussian was calculated from the SEPT2015 version of the Nearby Dwarf Galaxies catalogue¹⁵. The catalogue has entries for both velocity dispersions and rotation velocities, and for both gas and stars. To obtain a homogeneous estimate we use ‘effective’ dispersions, $\sigma_{\text{eff}} \equiv (\sigma^2 + 0.5v_{\text{rot}}^2)^{0.5}$, where v_{rot} is the inclination-corrected rotation velocity. When both gas and stellar kinematics are available, we use the highest value of σ_{eff} , as this typically is a rotation curve measurement from gas at large radii. Stellar masses were calculated directly from the V-band absolute magnitude assuming $M/L_V = 2.0$, for consistency with NGC1052–DF2. Five galaxies have a stellar mass that is within a factor of two of that of NGC 1052–DF2: these are IC 1613, NGC 6822, Sextans B and the M31 satellites NGC 147 and NGC 185. The average dispersion of these five galaxies is $\langle \sigma_{\text{eff}} \rangle = 32 \text{ km s}^{-1}$, with an r.m.s. variation of 8 km s^{-1} .

In Extended Data Fig. 2, we compare NGC 1052–DF2 to the nearby dwarf sample in the plane of velocity dispersion versus half-light radius, with the size of the symbols indicating the stellar mass. Comparing NGC1052–DF2 to other galaxies with velocity dispersions in this range, we find that its size is larger by a factor of about 10 and its stellar mass is larger by a factor of about 100.

Distance. The heliocentric radial velocity of NGC 1052–DF2 is $1,803 \pm 2 \text{ km s}^{-1}$, or $1,748 \pm 16 \text{ km s}^{-1}$ after correcting for the effects of the Virgo cluster, the Great Attractor and the Shapley supercluster on the local velocity field⁴⁴. For $H_0 = 70 \pm 3 \text{ km s}^{-1} \text{ Mpc}^{-1}$, a Hubble flow distance of $D_{\text{HF}} = 25 \pm 1 \text{ Mpc}$ is obtained. However, the proximity to NGC 1052 (14', or about 80 kpc in projection) strongly suggests that NGC1052–DF2 is associated with this massive elliptical galaxy. The distance to NGC 1052, as determined from surface brightness fluctuations (SBFs) and the fundamental plane^{9,10}, is $D_{1052} = 20.4 \pm 1.0 \text{ Mpc}$. The velocity of NGC1052–DF2 with respect to NGC 1052 is then $+293 \text{ km s}^{-1}$.

A third distance estimate can be obtained from the luminosity function of the compact objects. As discussed elsewhere (P.v.D. *et al.*, manuscript in preparation), the luminosity function has a narrow peak at $m_V \approx 22.0$. The canonical globular cluster luminosity function can be approximated by a Gaussian with a well-defined peak⁴⁵ at $M_V \approx -7.5$. If the compact objects are typical globular clusters, the implied distance is $D_{\text{GC}} \approx 8 \text{ Mpc}$. The main conclusions of the paper would be weakened considerably if the galaxy were so close to us. For this distance, the stellar mass estimate is an order of magnitude lower: $M_{\text{stars}} \approx 3 \times 10^7 M_{\odot}$. The four Local Group galaxies that have a stellar mass within a factor of two of this value (Fornax, Andromeda II, Andromeda VII and UGC 4879) have a mean dispersion of $\langle \sigma_{\text{eff}} \rangle = 11.7 \pm 0.5 \text{ km s}^{-1}$, only slightly higher than the upper limit to the dispersion of NGC1052–DF2. The peculiar velocity of the galaxy would be about $1,200 \text{ km s}^{-1}$; this is, of course, extreme, but it is difficult to argue that it is less likely than having a highly peculiar globular cluster population and a lack of dark matter.

Fortunately, we have independent information to verify the distance, namely the appearance of NGC1052–DF2 in the HST images. In Extended Data Fig. 3a, we show the central $33'' \times 33''$ of the galaxy in the HST I_{814} band. A smooth model of the galaxy, obtained by median filtering the image, was subtracted; background galaxies and globular clusters were masked. The mottled appearance is not due to noise but due to the variation in the number of giants contributing to each pixel. Following previous studies^{9,46,47}, we measure the SBF signal from this image and determine the distance from the SBF magnitude.

The azimuthally averaged power spectrum of the image is shown in Extended Data Fig. 3b. As is customary⁴⁶, the smallest and largest wavenumbers are omitted,

as they are dominated by residual large-scale structure in the image and by noise correlations, respectively. Again following previous studies^{46,47}, the power spectrum is fitted by a combination of a constant (dotted line) and the expectation power spectrum $E(k)$ (dashed line). The expectation power spectrum is the convolution of the power spectrum of the point spread function and that of the window function. The window function is the square root of the median-filtered model of the galaxy, multiplied by the mask containing the globular clusters and background galaxies.

The normalization of $E(k)$ is the SBF magnitude, m_{814} . We find $m_{814} = 29.45 \pm 0.10$. Using equation (2) in ref. 47, $V_{606} - I_{814} = 0.37 \pm 0.05$, and $g_{475} - I_{814} = 1.852(V_{606} - I_{814}) + 0.096$, the absolute SBF magnitude is $M_{814} = -1.94 \pm 0.17$. The uncertainty is a combination of the error in the $V_{606} - I_{814}$ colour and the systematic uncertainty in the extrapolation of the relation between $g_{475} - I_{814}$ colour and M_{814} (as determined from the difference between equations (1) and (2) in ref. 47). The SBF distance modulus $m - M = 31.39 \pm 0.20$, and the SBF distance $D_{\text{SBF}} = 19.0 \pm 1.7 \text{ Mpc}$. This result is consistent with D_{1052} and rules out the ‘globular cluster distance’ of $D_{\text{GC}} = 8 \text{ Mpc}$.

Stellar mass. We determined the stellar mass from a stellar population synthesis model⁴¹. A two-dimensional model galaxy was created using the ArtPop code⁴⁹ that matches the morphology, luminosity, colour and SBF signal of NGC1052–DF2. The model has a metallicity $[Z/H] = -1$ and an age of 11 Gyr. These parameters are consistent with the regular morphology of the galaxy and with spectroscopic constraints on the stellar populations of Coma cluster UDGs⁵⁰. For a Kroupa⁵¹ initial mass function, the stellar mass of this model is $1.8 \times 10^8 M_{\odot}$, similar to that obtained from a simple $M/L_V = 2.0$ conversion¹³ ($M_{\text{stars}} = 2.2 \times 10^8 M_{\odot}$). In the main text and below, we assume $M_{\text{stars}} \approx 2 \times 10^8 M_{\odot}$. We note that the uncertainty in the stellar mass is much smaller than that in the dynamical mass, for reasonable choices of the initial mass function.

Dynamical equilibrium. Some large low-surface-brightness objects are almost certainly in the process of disruption; examples are the ‘star pile’ in the galaxy cluster Abell 545^{52,53}, the boomerang-shaped galaxy DF4 in the field of M101⁵⁴, and Andromeda XIX in the Local Group⁵⁵ (marked in Extended Data Fig. 2). Andromeda XIX, and also And XXI and And XXV, are particularly informative as they combine large sizes with low velocity dispersions⁵⁶, and it has been suggested that tidal interactions have contributed to their unusual properties⁵⁵. (These galaxies are not direct analogues of NGC1052–DF2: the stellar masses of these Andromeda satellites are a factor of about 100 lower than that of NGC1052–DF2, and their dynamical M/L ratios are at least a factor of 10 higher.) In Extended Data Fig. 4c, we show the Gemini *i*-band image of NGC1052–DF2 (along with the Dragonfly and SDSS images). There is no convincing evidence of strong position angle twists or tidal features at least out to $R \approx 2R_e$ (see also the Dragonfly image in Extended Data Fig. 4a). The regular appearance strongly suggests that the object has survived in its present form for multiple dynamical times, and we infer that the kinematics can probably be interpreted in the context of a system that is in dynamical equilibrium.

We note that the regular morphology of NGC1052–DF2 also provides an interesting constraint on its formation time: the orbital velocity in the outer parts is 5 Gyr, which means it must have formed very early in order to lose any sign of its assembly. Furthermore, it provides a lower limit for the 3D distance between NGC1052–DF2 and the massive elliptical galaxy NGC 1052. The Jacobi radius (that is, the distance from the centre of the galaxy to the first Lagrangian point) is given by⁵⁷

$$R_J = \left(\frac{GM}{2V_{1052}^2} \right)^{1/3} R_{1052}^{2/3} \quad (1)$$

with R_{1052} the distance between NGC1052–DF2 and NGC 1052, and V_{1052} the circular velocity of NGC 1052. Taking $R_J > 5 \text{ kpc}$, $M \approx 2 \times 10^8 M_{\odot}$ and $V_{1052} \approx 300 \text{ km s}^{-1}$ (the velocity difference between the two galaxies, as well as approximately $\sqrt{2}\sigma_{1052}$), we obtain R_{1052} greater than about 160 kpc, a factor of two larger than the projected distance.

Source of dynamical support. The morphology of the galaxy strongly indicates that it is supported by random motions rather than rotation: the Sérsic index is 0.6, similar to that of dSph galaxies; the isophotes are elliptical rather than disk-like; and there are no bars, spiral arms, or other features that might be expected in a thin disk. The galaxy has not been detected in moderately deep H I observations⁵⁸. It is also difficult to imagine a physical model for the formation of a huge, extremely thin disk of massive blue globular clusters, even in spiral galaxies: although the kinematics of the metal-rich globular cluster population in M31 are clearly related to its disk, the metal-poor ones have a large velocity dispersion⁵⁹. Finally, there is no evidence for a velocity gradient. In Extended Data Fig. 5a, we show the measured velocities of the globular clusters as a function of the projected distance along the major axis. There is no coherent pattern. Based on these arguments, our default mass measurement assumes that the galaxy is supported by random motions.

Dynamical mass measurement. Following a previous study of the kinematics of globular clusters in a UDG¹⁴, we use the TME to determine the dynamical mass. This method was developed to determine the enclosed mass from an ensemble of discrete tracers, such as satellite galaxies or globular clusters^{16,48}. The mass within the distance of the outermost object is given by

$$M_{\text{TME}} = \frac{C}{G} \langle v^2 r^\alpha \rangle r_{\text{out}}^{1-\alpha} \quad (2)$$

with v the velocities of individual clusters with respect to the mean, r the projected distances of the clusters from the centre of the galaxy, r_{out} the distance of the furthest cluster, and α the slope of the potential (with density $\rho \propto r^{-(\alpha+2)}$). For the case $\alpha = 1$, the potential is similar to that of a point mass; $\alpha = 0$ corresponds to $\rho \propto r^{-2}$ and a flat rotation curve; and for $\alpha = -1$ the density $\rho \propto r^{-1}$. Equation (2) does not take observational errors or outliers into account; we therefore introduce the modified expression

$$M_{\text{TME}} \approx \frac{C}{G} [S_{\text{bi}}(v_{\text{intr}} r^{\alpha/2})]^2 r_{\text{out}}^{1-\alpha} \quad (3)$$

Here observational errors are taken into account by setting $v_{\text{intr}} = f^{-1}(\Delta v_{\text{obs}})$, with Δv_{obs} listed in Fig. 2 and $f^{-1} = \sigma_{\text{intr}}/\sigma_{\text{obs}}$. $S_{\text{bi}}(x)$ denotes the biweight estimator of the width of the distribution. Note that equation (3) reduces to $M_{\text{TME}} = (C/G)\sigma_{\text{intr}}^2 r_{\text{out}}$, for $\alpha = 0$. The constant C is given by

$$C = \frac{4\Gamma(\frac{\alpha}{2} + \frac{5}{2})}{\sqrt{\pi}\Gamma(\frac{\alpha}{2} + 1)} \frac{\alpha + \gamma + 1 - 2\beta}{\alpha + 3 - \beta(\alpha + 2)} \quad (4)$$

with γ the power-law slope of the density profile of the clusters, $\beta = 1 - \sigma_t^2/\sigma_r^2$ the Binney anisotropy parameter and $\Gamma(x)$ the gamma function. We determine the 3D density profile from a power-law fit to the distribution of the spectroscopically confirmed globular clusters, finding $\gamma = 0.9 \pm 0.3$.

For an isothermal velocity dispersion profile ($\alpha = 0$) and isotropic orbits ($\beta = 0$) we determine $M_{\text{TME}} < 3.4 \times 10^8 M_\odot$ at 90% confidence. The results are not very sensitive to the assumed slope of the potential or moderate anisotropy. Changing α to 1 or -1 reduces the mass by 10% or 20% respectively. Tangential anisotropy with $\sigma_t^2 = 2\sigma_r^2$ increases the mass limit to $M_{\text{TME}} < 4.2 \times 10^8 M_\odot$; radial anisotropy with $\sigma_r^2 = 0.5\sigma_t^2$ yields $M_{\text{TME}} < 2.4 \times 10^8 M_\odot$. We also consider errors in the density profile of the globular clusters; for $\gamma = 0.5$ the mass decreases by 20% and for $\gamma = 1.5$ the mass increases by 30%.

Robustness tests. As a test of the robustness of our results, we consider three alternative mass estimates. The first is the dynamical mass within the half-number radius of the globular cluster system¹⁷. This mass estimate does not extend as far in radius as the TME method but is less sensitive to the assumed level of anisotropy. For $R_{\text{gc}} = 3.1$ kpc and $\sigma_{\text{intr}} < 10.5$ km s⁻¹ we find $M_{\text{dyn}} < 3.2 \times 10^8 M_\odot$ (see Fig. 4). As the halo profile is still rising at $R = 3.1$ kpc, the constraint on the halo mass is weaker than our default value, and we find $M_{\text{halo}} < 8 \times 10^8 M_\odot$.

The second test replaces σ_{bi} with σ_{rms} , even though the r.m.s. is driven by a single cluster (GC-98) and the velocity distribution of the other nine objects is inconsistent with this. The observed r.m.s. is $\sigma_{\text{obs,rms}} = 14.3$ km s⁻¹, or $\sigma_{\text{intr,rms}} = 12.2$ km s⁻¹ after taking observational errors into account. The implied TME mass is $M_{\text{dyn}} \approx 5 \times 10^8 M_\odot$, and the halo mass $M_{\text{halo}} \approx 3 \times 10^8 M_\odot$.

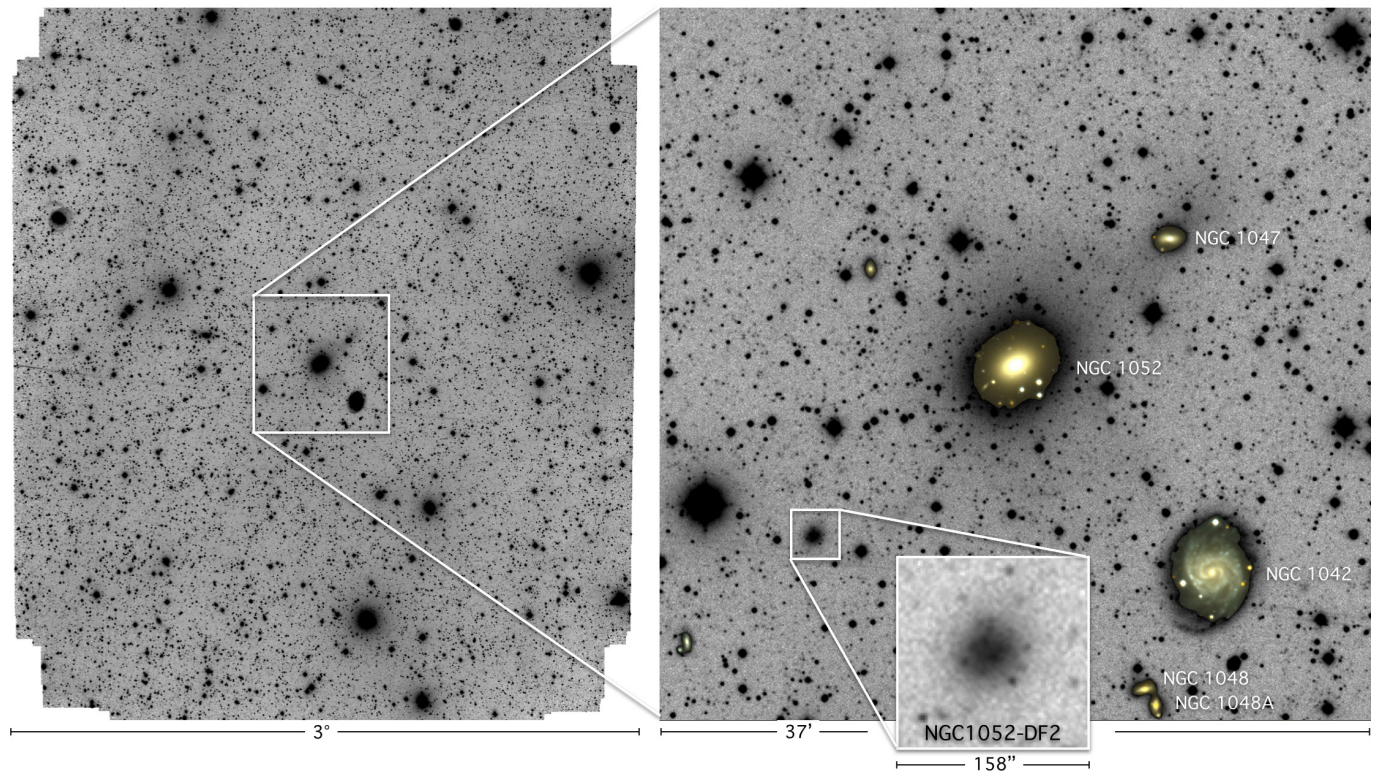
The third test sets the arguments against a disk aside and assumes that the observed velocities reflect rotation in an inclined, infinitely thin disk. The axis ratio of NGC1052-DF2 is $b/a = 0.85 \pm 0.02$, which means that the inclination-corrected velocities are a factor of $(\sin(\cos^{-1}(b/a)))^{-1} \approx 1.9$ higher than the observed ones. Assuming an (unphysical) disk dispersion of 0 km s⁻¹, the inclination-corrected rotation velocity would be $v_{\text{rot}} \approx 1.4 \times 3.2 \times 1.9 = 9.1^{+1.4}_{-0.9}$ km s⁻¹, where it is assumed that the rotation velocity is approximately 1.4 times the line-of-sight velocity dispersion^{60,61}. The enclosed mass within $R = 7.6$ kpc would be $M_{\text{disk}} = 1.4^{+7.6}_{-1.4} \times 10^8 M_\odot$.

For all these mass estimates, the implied ratio $M_{\text{halo}}/M_{\text{stars}}$ is less than 4. This is the lowest ratio measured for any galaxy and two orders of magnitude below the canonical relation between stellar mass and halo mass.

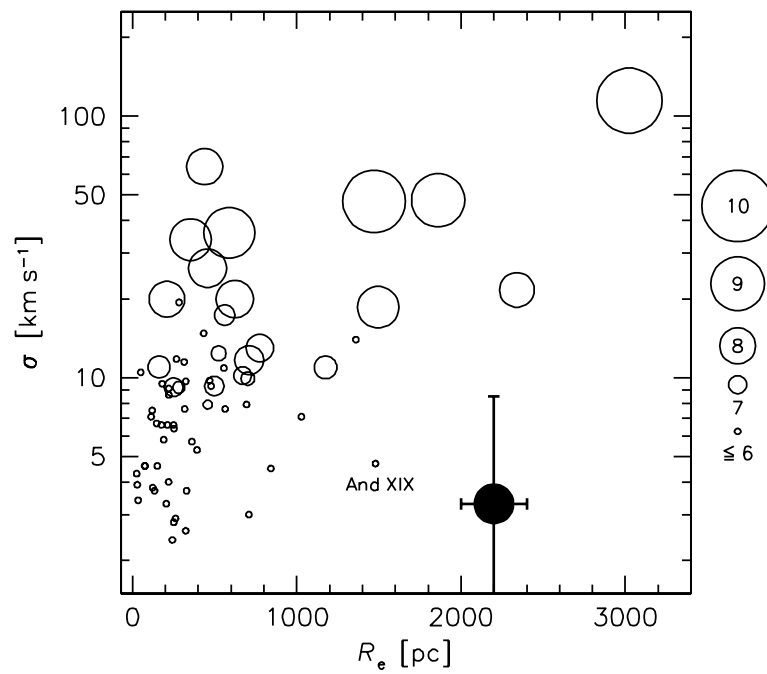
Data availability. The HST data are available in the Mikulski Archive for Space Telescopes (MAST; <http://archive.stsci.edu>), under programme ID 14644. All other data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability. We have made use of standard data reduction tools in the IRAF and Python environments, and the publicly available codes SExtractor³⁵, GALFIT¹¹ and emcee⁴⁰.

31. Merritt, A., van Dokkum, P., Abraham, R. & Zhang, J. The Dragonfly Nearby Galaxies Survey. I. Substantial variation in the diffuse stellar halos around spiral galaxies. *Astrophys. J.* **830**, 62 (2016).
32. Abolfathi, B. et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: first spectroscopic data from the extended Baryon Oscillation Sky Survey and from the second phase of the Apache Point Observatory Galactic Evolution Experiment. Preprint available at <http://arxiv.org/abs/1707.09322> (2017).
33. Napolitano, N. R. et al. The Planetary Nebula Spectrograph elliptical galaxy survey: the dark matter in NGC 4494. *Mon. Not. R. Astron. Soc.* **393**, 329–353 (2009).
34. Hook, I. M. et al. The Gemini-North Multi-Object Spectrograph: Performance in imaging, long-slit, and multi-object spectroscopic modes. *Publ. Astron. Soc. Pac.* **116**, 425–440 (2004).
35. Bertin, E. & Arnouts, S. SExtractor: Software for source extraction. *Astron. Astrophys. Suppl.* **117**, 393–404 (1996).
36. Sersic, J. L. *Atlas de galaxias australes* (Observatorio Astronómico, Córdoba, 1968).
37. Faber, S. M. et al. in *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*, Proc. SPIE Vol. 4841 (eds Iye, M. & Moorwood, A. F. M.) 1657–1669 (SPIE, 2003).
38. Oke, J. B. et al. The Keck Low-Resolution Imaging Spectrometer. *Publ. Astron. Soc. Pac.* **107**, 375 (1995).
39. van Dokkum, P. et al. A high stellar velocity dispersion and ~ 100 globular clusters for the ultra-diffuse galaxy Dragonfly 44. *Astrophys. J.* **828**, L6 (2016).
40. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: The MCMC hammer. *Publ. Astron. Soc. Pac.* **125**, 306–312 (2013).
41. Conroy, C., Gunn, J. E. & White, M. The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *Astrophys. J.* **699**, 486–506 (2009).
42. van Dokkum, P. G., Kriek, M. & Franx, M. A high stellar velocity dispersion for a compact massive galaxy at redshift $z = 2.186$. *Nature* **460**, 717–719 (2009).
43. Beers, T. C., Flynn, K. & Gebhardt, K. Measures of location and scale for velocities in clusters of galaxies—a robust approach. *Astron. J.* **100**, 32–46 (1990).
44. Mould, J. R. et al. The Hubble Space Telescope Key Project on the extragalactic distance scale. XXVIII. Combining the constraints on the Hubble constant. *Astrophys. J.* **529**, 786–794 (2000).
45. Rejkuba, M. Globular cluster luminosity function as distance indicator. *Astrophys. Space Sci.* **341**, 195–206 (2012).
46. Mei, S. et al. The Advanced Camera for Surveys Virgo Cluster Survey. V. Surface brightness fluctuation calibration for giant and dwarf early-type galaxies. *Astrophys. J.* **625**, 121–129 (2005).
47. Blakeslee, J. P. et al. Surface brightness fluctuations in the Hubble Space Telescope ACS/WFC F814W bandpass and an update on galaxy distances. *Astrophys. J.* **724**, 657–668 (2010).
48. Bahcall, J. N. & Tremaine, S. Methods for determining the masses of spherical systems. I.—Test particles around a point mass. *Astrophys. J.* **244**, 805–819 (1981).
49. Danieli, S., van Dokkum, P. & Conroy, C. Hunting faint dwarf galaxies in the field using integrated light surveys. Preprint at <http://arxiv.org/abs/1711.00860> (2017).
50. Gu, M. et al. Low metallicities and old ages for three ultra-diffuse galaxies in the Coma cluster. Preprint at <http://arxiv.org/abs/1709.07003> (2017).
51. Kroupa, P. On the variation of the initial mass function. *Mon. Not. R. Astron. Soc.* **322**, 231–246 (2001).
52. Struble, M. F. Optical discovery of intracluster matter on the Palomar Observatory Sky Survey—the star pile in A545. *Astrophys. J.* **330**, L25–L28 (1988).
53. Salinas, R. et al. Crazy heart: kinematics of the “star pile” in Abell 545. *Astron. Astrophys.* **528**, A61 (2011).
54. Merritt, A. et al. The Dragonfly Nearby Galaxies Survey. II. Ultra-diffuse galaxies near the elliptical galaxy NGC 5485. *Astrophys. J.* **833**, 168 (2016).
55. Collins, M. L. M. et al. A kinematic study of the Andromeda dwarf spheroidal system. *Astrophys. J.* **768**, 172 (2013).
56. Tollerud, E. J. et al. The SPLASH Survey: spectroscopy of 15 M31 dwarf spheroidal satellite galaxies. *Astrophys. J.* **752**, 45 (2012).
57. Baumgardt, H., Parmentier, G., Gieles, M. & Vesperini, E. Evidence for two populations of Galactic globular clusters from the ratio of their half-mass to Jacobi radii. *Mon. Not. R. Astron. Soc.* **401**, 1832–1838 (2010).
58. McKay, N. P. F. et al. The discovery of new galaxy members in the NGC 5044 and 1052 groups. *Mon. Not. R. Astron. Soc.* **352**, 1121–1134 (2004).
59. Caldwell, N. & Romanowsky, A. J. Star clusters in M31. VII. Global kinematics and metallicity subpopulations of the globular clusters. *Astrophys. J.* **824**, 42 (2016).
60. Franx, M. *Galactic Bulges*, Proc. 153th Symp. International Astronomical Union (ed. Dejonghe, H. & Habing, H. J.) 243–262 (IAU, 1993).
61. Kochanek, C. S. The dynamics of luminous galaxies in isothermal halos. *Astrophys. J.* **436**, 56–66 (1994).

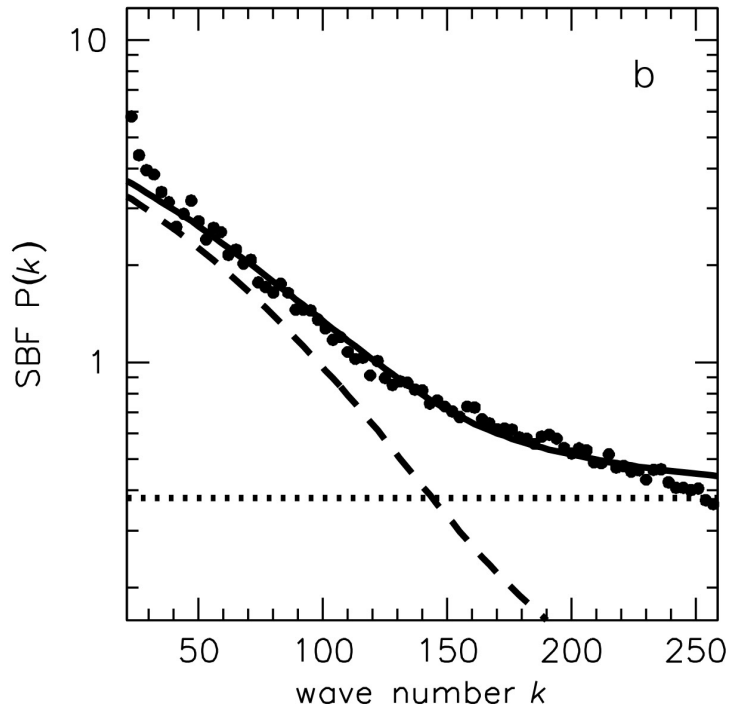
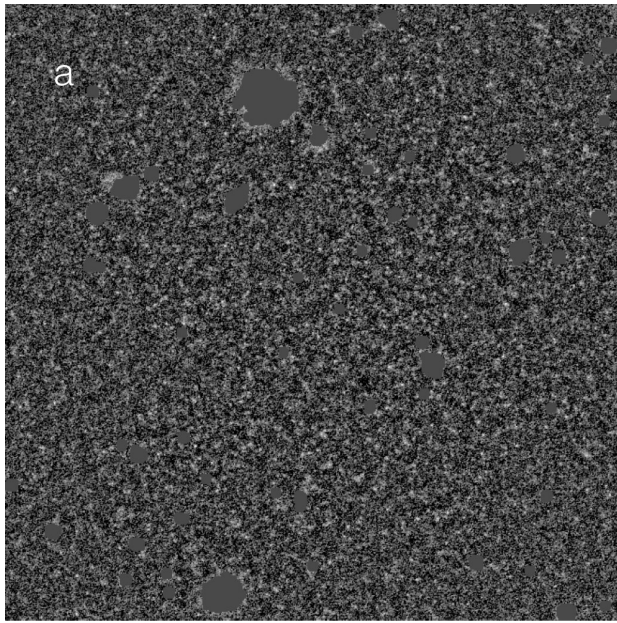


Extended Data Figure 1 | NGC1052-DF2 in the Dragonfly field. The full Dragonfly field, approximately 11 degree^2 , centred on NGC 1052. The zoom-in shows the immediate surroundings of NGC 1052, with NGC1052-DF2 highlighted in the inset.



Extended Data Figure 2 | Comparison to Local Group galaxies. Open symbols are galaxies from the Nearby Dwarf Galaxies catalogue¹⁵ and the solid symbol with error bars is NGC1052-DF2. The size of each symbol indicates the logarithm of the stellar mass, as shown in the key. There are no galaxies in the Local Group that are similar to NGC1052-DF2. Galaxies

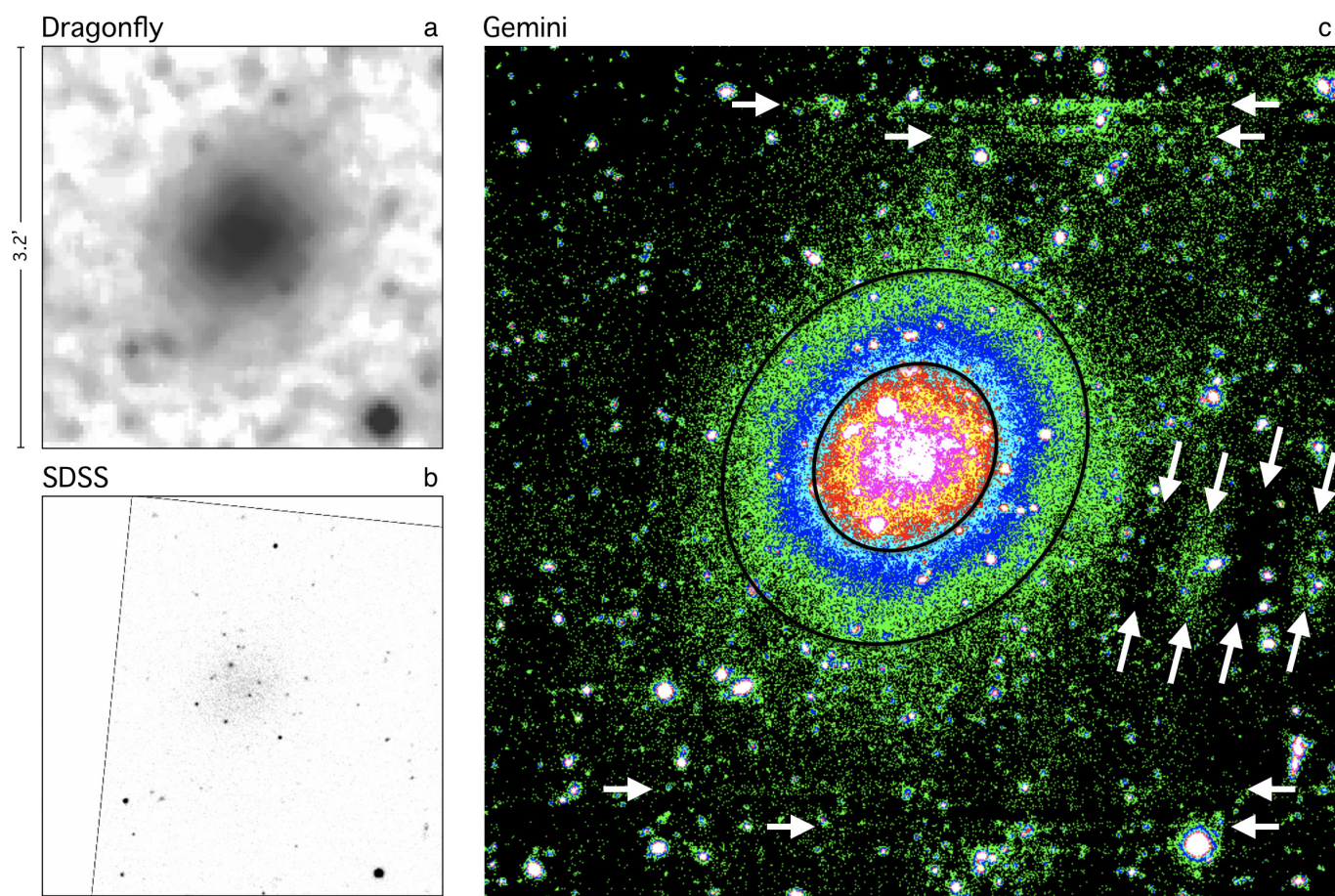
with a similar velocity dispersion are a factor of about 10 smaller and have stellar masses that are a factor of about 100 larger. The object labelled And XIX is an Andromeda satellite that is thought to be in the process of tidal disruption⁵⁵.



Extended Data Figure 3 | Analysis of surface brightness fluctuation.

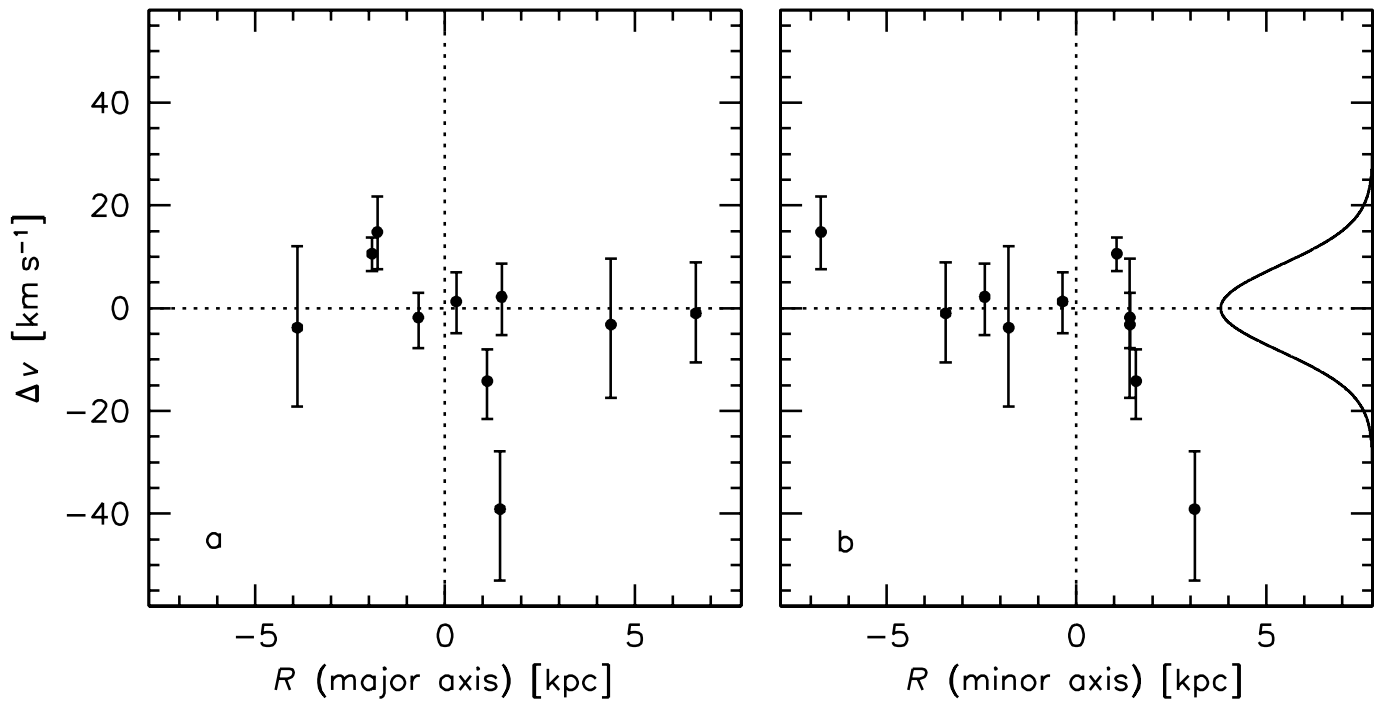
We use the surface brightness fluctuation (SBF) signal in the HST I_{814} band to constrain the distance to NGC1052–DF2. **a**, The galaxy after subtracting a smooth model and masking background galaxies and globular clusters. The image spans $33'' \times 33''$. **b**, The azimuthally averaged power spectrum. Following previous studies^{9,46,47}, the power spectrum

is fitted by a combination of a constant (dotted line) and an expectation power spectrum $E(k)$ (dashed line). From the normalization of $E(k)$ we find that the SBF magnitude $m_{814} = 29.45 \pm 0.10$. The implied distance is $D_{\text{SBF}} = 19.0 \pm 1.7$ Mpc, consistent with the 20 Mpc distance of the luminous elliptical galaxy NGC 1052.



Extended Data Figure 4 | Morphological coherence. **a**, The sum of g and r images taken with the Dragonfly Telephoto Array. The image was smoothed by a $10'' \times 10''$ median filter to bring out faint emission. The lowest surface brightness levels visible in the image are about $29 \text{ mag arcsec}^{-2}$. **b**, Sum of SDSS g , r and i images. In SDSS, the overdensity of compact objects stands out. **c**, The Gemini-North i -band image of

NGC1052-DF2, which provides the best information on the morphology of the galaxy. Black ellipses mark $R = R_e$ and $R = 2R_e$. White arrows mark the most obvious reduction artefacts. The galaxy is regular out to at least $R \approx 2R_e$, with a well-defined centre and a position angle and axis ratio that do not vary strongly with radius.



Extended Data Figure 5 | Are the globular clusters in a thin rotating disk? **a, b**, Globular cluster velocities as a function of projected position along the major axis (**a**) and the minor axis (**b**). There is no evidence for any trends. For reference, a Gaussian with $\sigma = 8.4 \text{ km s}^{-1}$ is shown in **b**.

A programmable two-qubit quantum processor in silicon

T. F. Watson¹, S. G. J. Philips¹, E. Kawakami¹, D. R. Ward², P. Scarlino¹, M. Veldhorst¹, D. E. Savage², M. G. Lagally², Mark Friesen², S. N. Coppersmith², M. A. Eriksson² & L. M. K. Vandersypen¹

Now that it is possible to achieve measurement and control fidelities for individual quantum bits (qubits) above the threshold for fault tolerance, attention is moving towards the difficult task of scaling up the number of physical qubits to the large numbers that are needed for fault-tolerant quantum computing^{1,2}. In this context, quantum-dot-based spin qubits could have substantial advantages over other types of qubit owing to their potential for all-electrical operation and ability to be integrated at high density onto an industrial platform^{3–5}. Initialization, readout and single- and two-qubit gates have been demonstrated in various quantum-dot-based qubit representations^{6–9}. However, as seen with small-scale demonstrations of quantum computers using other types of qubit^{10–13}, combining these elements leads to challenges related to qubit crosstalk, state leakage, calibration and control hardware. Here we overcome these challenges by using carefully designed control techniques to demonstrate a programmable two-qubit quantum processor in a silicon device that can perform the Deutsch–Jozsa algorithm and the Grover search algorithm—canonical examples of quantum algorithms that outperform their classical analogues. We characterize the entanglement in our processor by using quantum-state tomography of Bell states, measuring state fidelities of 85–89 per cent and concurrences of 73–82 per cent. These results pave the way for larger-scale quantum computers that use spins confined to quantum dots.

Solid-state approaches to quantum computing are challenging to realize owing to unwanted interactions between the qubit and the host material. For quantum-dot-based qubits, charge and nuclear-spin noise are the dominant sources of decoherence and gate errors. Some of these effects can be cancelled out by using dynamical decoupling¹⁴ or decoherence-free subspaces^{9,15}, but there has also been substantial progress in reducing these noise sources by growing better oxides and heterostructures¹⁶ and moving to silicon, owing to its naturally low abundance of nuclear-spin isotopes that can be removed through isotopic purification¹⁷. These material developments have greatly extended the coherence times of qubits, enabling single-qubit gate fidelities of above 99%^{18–21}, and recently resulted in the demonstration of a controlled-phase (CZ) gate between two single-electron-spin qubits in a silicon metal–oxide–semiconductor (Si-MOS) device⁸. Here, we show that with two single-electron-spin qubits in a natural silicon/silicon germanium (Si/SiGe) double quantum dot we can combine initialization, readout, single- and two-qubit gates to form a programmable quantum processor in silicon that can perform simple quantum algorithms.

A schematic of the two-qubit quantum processor is shown in Fig. 1a. The device is similar to that described previously²² except for an additional micrometre-scale magnet ('micromagnet'). A two-dimensional electron gas is formed in the natural silicon quantum well of a SiGe heterostructure using two accumulation gates. The double quantum dot is defined in this two-dimensional electron gas by applying negative voltages to the depletion gates, with the estimated positions of the first (D1) and second (D2) quantum dot shown by the purple and orange circles, respectively. The two qubits, Q1 and Q2, are defined by applying

a finite magnetic field $B_{\text{ext}} = 617$ mT and using the Zeeman-split spin-down $|0\rangle$ and spin-up $|1\rangle$ states of single electrons, respectively, which are confined in D1 and D2. The initialization and readout of Q2 are performed by spin-selective tunnelling to a reservoir²³, while Q1 is initialized at a spin-relaxation hotspot²⁴ and measured via Q2 using a controlled-rotation (CROT) gate. The complete measurement sequence and set-up are described in Extended Data Figs 1 and 2. We achieve initialization ('I' subscript) and readout ('m' subscript) fidelities for the two qubits of $F_{I1} > 99\%$, $F_{I2} > 99\%$, $F_{m1} = 73\%$ and $F_{m2} = 81\%$ (see Methods).

The coherent individual control of both qubits is achieved by patterning three cobalt micromagnets on top of the device (Fig. 1a). These micromagnets provide a magnetic-field gradient with a component that is perpendicular to the external magnetic field for electric dipole spin resonance (EDSR)²⁵. Furthermore, the field gradient across the two dots results in qubit frequencies that are well separated ($f_{Q1} = 18.4$ GHz, $f_{Q2} = 19.7$ GHz), which allows the qubits to be addressed independently. For both qubits, we achieve Rabi frequencies of $f_R = \omega_R/(2\pi) = 2$ MHz and perform single-qubit X and Y gates by using vector modulation of the microwave drive signals. Here, we define an X (Y) gate to be a $\pi/2$ rotation around \hat{x} (\hat{y}), and henceforth define a π rotation to be X^2 (Y^2). We measure the spin relaxation time (T_1), free evolution time (T_2^*) and the Hahn echo decay time ($T_{2\text{Hahn}}$) of Q1 in the (1, 1) regime (where (m , n) denotes a configuration with m electrons in D1 and n electrons in D2) to be $T_1 > 50$ ms, $T_2^* = 1.0 \pm 0.1$ μ s and $T_{2\text{Hahn}} = 19 \pm 3$ μ s, and those of Q2 to be $T_1 > 3.7 \pm 0.5$ ms, $T_2^* = 0.6 \pm 0.1$ μ s and $T_{2\text{Hahn}} = 7 \pm 1$ μ s (Extended Data Fig. 3). Using randomized benchmarking^{20,26} we find average single-qubit gate fidelities of 98.8% for Q1 and 98.0% for Q2 (Extended Data Fig. 4)—close to the fault-tolerance error threshold for surface codes²⁷.

Universal quantum computing requires the implementation of single- and two-qubit gates. In our quantum processor we implement a two-qubit CZ gate^{8,28}. This gate can be understood by considering the energy-level diagram for two electron spins in a double quantum dot (Fig. 1b), in the regime in which the Zeeman-energy difference is comparable to the inter-dot tunnel coupling ($\Delta E_Z \approx t_c$). In Fig. 1b we plot the energies of the two-spin states ($|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$) in the (1, 1) charge regime and the singlet ground state $S(0, 2)$ in the (0, 2) charge regime as a function of the detuning ϵ . Here, detuning describes the energy difference between the (1, 1) and (0, 2) charge states of the double quantum dot, controlled by the voltage applied to plunger gate P1 (Extended Data Fig. 2). The anticrossing between the $S(0, 2)$ and the antiparallel $|01\rangle$ and $|10\rangle$ states causes the energy of the antiparallel states to decrease by $J(\epsilon)/2$ as the detuning is decreased (Fig. 1b), where $J(\epsilon)$ is the exchange coupling between the two electron spins.

The energy structure of the two-electron system can be probed by performing microwave spectroscopy as a function of detuning (Fig. 1c). At negative detuning, the resonance frequency (Zeeman energy) increases linearly (dashed line) owing to the electron

¹QuTech and the Kavli Institute of Nanoscience, Delft University of Technology, 2600 GA Delft, The Netherlands. ²University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

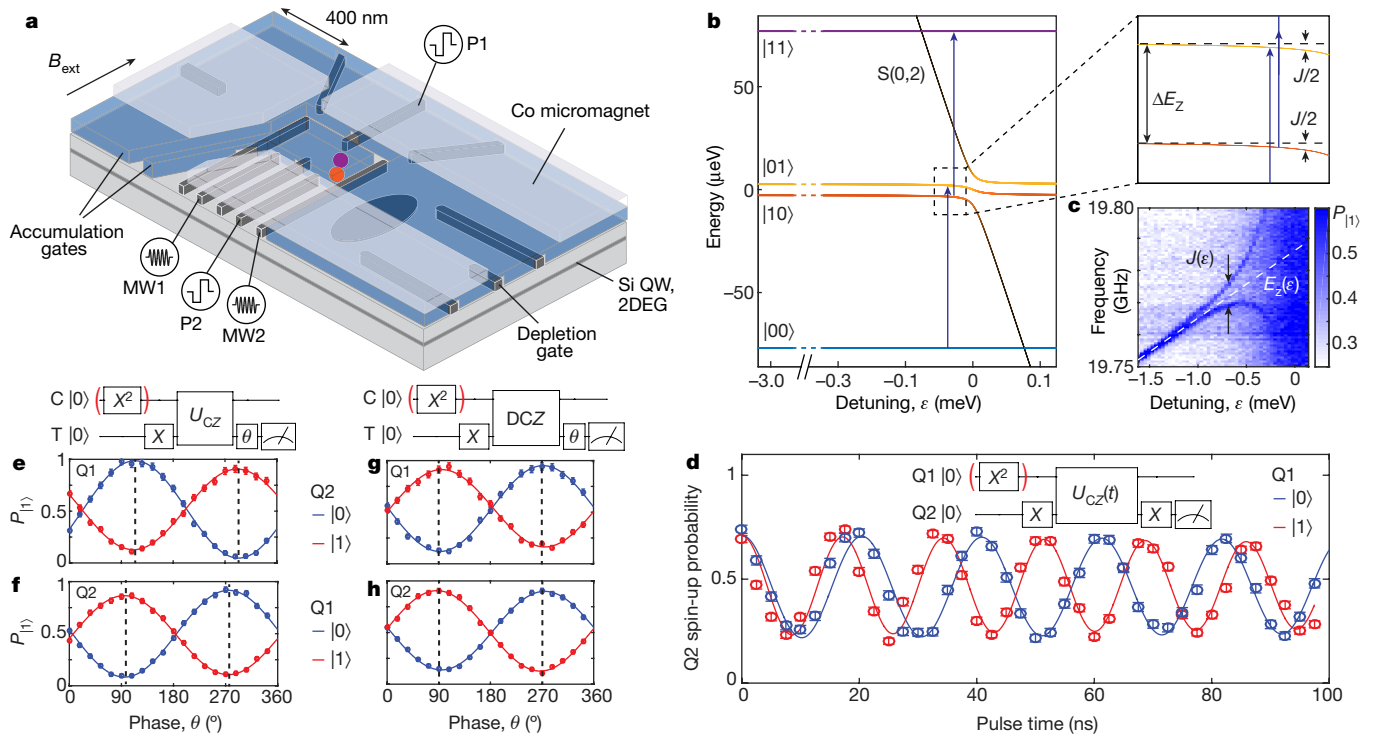


Figure 1 | Two-qubit quantum processor in silicon. **a**, Schematic of a Si/SiGe double-quantum-dot device, showing the estimated positions of quantum dots D1 (purple circle) and D2 (orange circle) that are used to confine two electron-spin qubits Q1 and Q2, respectively. Both quantum dots were formed from the two-dimensional electron gas (2DEG) in the silicon quantum well (Si QW) on the right side of the device to achieve an inter-dot tunnel coupling that is suitable for two-qubit gates. The positions of the dots were realized by tuning the numerous accumulation and depletion gates, but were probably helped by disorder in the Si/SiGe heterostructure. The ellipse shows the position of the quantum-dot sensor used for spin readout. Microwave signals MW1 and MW2 are used to perform electric dipole spin resonance (EDSR), mediated by the cobalt (Co) micromagnets, on Q1 and Q2, respectively, while voltage pulses are applied to plunger gates P1 and P2 for qubit manipulation and readout. **b**, Energy-level diagram of two electron spins in a double quantum dot as a function of the detuning energy ε between the $(1, 1)$ and $(0, 2)$ charge states. Towards zero detuning the energy levels of the anti-parallel spin states shift by half the exchange energy J (see inset). **c**, Microwave spectroscopy of Q2 showing the spin-up probability P_{11} versus the MW2 frequency and detuning energy after initialization of Q1 to $(|0\rangle + |1\rangle)/\sqrt{2}$. The detuning voltage was converted to energy using a lever arm of $\alpha = 0.09e$, where e is the electron charge (Extended Data Fig. 5). The map

wavefunction moving in the magnetic-field gradient. At more positive detuning, closer to the $(0, 2)$ regime, the exchange energy is substantial compared to the line width of the resonance $J/h > \omega_R$ (where $h = 2\pi\hbar$ is the Planck constant), resulting in two clear resonances. Applying a π pulse at one of these frequencies results in a CROT gate, which is used to perform the projective measurement of Q1 via the readout of Q2 (Extended Data Fig. 6).

The CZ gate is implemented by applying a detuning pulse for a fixed amount of time t , which shifts the energy of the antiparallel states. Throughout the pulse, we stay in the regime in which $J(\varepsilon) \ll \Delta E_Z$, so the energy eigenstates of the system are still the two-spin product states and the two-qubit interaction can be approximated by an Ising Hamiltonian, leading to the following unitary operation:

$$U_{CZ}(t) = Z_1(\theta_1)Z_2(\theta_2) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{ij(\varepsilon)t/(2h)} & 0 & 0 \\ 0 & 0 & e^{ij(\varepsilon)t/(2h)} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

shows that Q2 has two different resonant frequencies (blue arrows in **b**) depending on the spin state of Q1, which are separated by the exchange energy J . **d**, The spin-up probability of Q2 after applying the Ramsey sequence (see inset) in which the duration of the detuning pulse is varied between two X gates on Q2, and the control Q1 is initialized to spin-down (blue curve) or spin-up (red curve). **e**, The spin-up probability of the target qubit (T; Q1) after applying the Ramsey sequence shown above the plot, in which a CZ gate is applied between two $\pi/2$ pulses and the phase of the second $\pi/2$ pulse is varied. Here, the control (C) qubit (C; Q2) is initialized to spin-down (blue curve) or spin-up (red curve) and the spin-up probability has been normalized to remove initialization and readout errors. The exchange energy during the CZ gate is $J/h = 10$ MHz. **f**, Similar to **e**, but with Q2 as the target qubit and Q1 as the control qubit. In **e** and **f**, the black dashed lines show the \hat{z} rotations on Q1 and Q2 that are needed to form the CZ_{ij} gates. **g**, **h**, Similar to **e** and **f**, but using a decoupled version of the CZ gate (DCZ gate), which removes the unconditional \hat{z} rotations due to the detuning dependence of $E_Z(\varepsilon)$. Consequently, the \hat{z} rotations that are required to form the CZ_{ij} gates (dashed black lines) are always at 90° and 270° , which simplifies the calibration. All error bars are 1σ from the mean, calculated using a Monte Carlo method (see Methods).

where the basis states are $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$, and $Z_1(\theta_1)$ and $Z_2(\theta_2)$ are rotations around \hat{z} caused by the change in the Zeeman energy of the qubits due to the magnetic-field gradient. The CZ gate is advantageous over the CROT gate because it is faster and less time is spent at low detuning, at which the qubits are more sensitive to charge noise. In addition, we observed that performing the CROT gate with EDSR can lead to state leakage into the $S(0, 2)$ state, seen in Fig. 1c by the increase in background dark counts near $\varepsilon = 0$. The CZ gate is demonstrated in Fig. 1d; we vary the duration of a CZ voltage pulse between two X gates on Q2 in a Ramsey experiment, showing that the frequency of the \hat{z} rotation on Q2 is conditional on the spin state of Q1. The processor's primitive two-qubit gates, $CZ_{ij}|m, n\rangle = (-1)^{\delta(i,m)\delta(j,n)}|m, n\rangle$ for $i, j, m, n \in \{0, 1\}$, are constructed by applying the CZ gate for a time $t = \pi\hbar/J$ followed by \hat{z} rotations on Q1 and Q2, $CZ_{ij} = Z_1[(-1)^{j\pi/2} - \theta_1]Z_2[(-1)^{i\pi/2} - \theta_2]U_{CZ}(\pi\hbar/J)$. Rather than physically performing the \hat{z} rotations, we change the reference frame in the software by incorporating the rotation angles θ_1 and θ_2 into the phase of any subsequent microwave pulses¹⁰.

Combining single- and two-qubit gates together with initialization and readout, we demonstrate a programmable processor—whereby we can program arbitrary sequences for the two-qubit chip to execute within the coherence times of the qubits. To achieve this, several challenges needed to be overcome. The device had to be tuned further so that during single-qubit gates the exchange coupling was low ($J_{\text{off}}/\hbar = 0.27$ MHz; Extended Data Fig. 7) compared to our single-qubit (about 2 MHz) and two-qubit (about 6–10 MHz) gate times. Tuning was also required to raise the energy of low-lying valley-excited states to prevent them from being populated during initialization²². Furthermore, we observed that applying microwave pulses on Q1 shifts the resonance frequency of Q2 by around 2 MHz. We rule out the AC Stark shift and effects from coupling between the spins and from heating as possible explanations, but find that the properties of the quantum dots affect the frequency shift (Supplementary Information, section S1). Although the origin of the shift is unknown, we keep the resonance frequency of Q2 fixed during single-qubit gates by applying an off-resonant pulse (30 MHz) to Q1 if Q1 is idle.

Before running sequences on the quantum processor, all gates need to be properly calibrated. The single-qubit X and Y gates were calibrated using both a Ramsey sequence and the AllXY calibration sequence to determine the qubit resonance frequency and the power needed to perform a $\pi/2$ gate (Supplementary Information, section S2). To calibrate the CZ_{ij} gates we performed the Ramsey sequence shown in Fig. 1e and varied the phase of the second $\pi/2$ gate. In Fig. 1e we show the results of this measurement, for which Q1 is the target qubit and the control qubit Q2 is prepared in either the $|0\rangle$ (blue curve) or $|1\rangle$ (red curve) state. The duration of the CZ gate is calibrated so that the blue and red curves are 180° out of phase. These measurements also determine the \hat{z} rotation on Q1 that is needed to form CZ_{ij} , which corresponds to the phase of the second $\pi/2$ gate, which either maximizes or minimizes the spin-up probability for Q2 (dashed lines in Fig. 1e). The \hat{z} rotation needed for Q2 is calibrated by performing a similar measurement, but with the roles of Q1 and Q2 switched (Fig. 1f).

The \hat{z} rotations in equation (1) can be eliminated by using a decoupled CZ gate, $\text{DCZ} = U_{\text{CZ}}(\pi\hbar/(2J))X_1^2X_2^2U_{\text{CZ}}(\pi\hbar/(2J))$, which incorporates refocusing pulses and can be used to perform $\text{DCZ}_{ij} = X_1^2X_2^2\text{CZ}_{ij} = Z_1[(-1)^j\pi/2]Z_2[(-1)^i\pi/2]\text{DCZ}$. This operation is demonstrated in the Ramsey experiment shown in Fig. 1g, h, in which the minimum and maximum spin-up probabilities occur at a phase of either 90° or 270° . In addition to removing the need to calibrate the required \hat{z} rotations, the DCZ gate is advantageous because it cancels out the effect of the low-frequency noise that couples to the spins via $\sigma_Z \otimes I$ and $I \otimes \sigma_Z$ terms during the implementation of the gate, where I is the 2×2 identity matrix and σ_Z is the Pauli z matrix.

After proper calibration, we characterize entanglement in our quantum processor by preparing Bell states and reconstructing the two-qubit density matrix using quantum-state tomography. The quantum circuit for the experiment is shown in Fig. 2a. The Bell states are prepared using a combination of single-qubit gates and the decoupled two-qubit DCZ_{ij} gates. The density matrix is reconstructed by measuring two-spin probabilities for the nine combinations of three different measurement bases (x, y, z) with 10,000 repetitions (Methods). In our readout scheme the states are projected into the z basis, while measurements in the other bases are achieved by performing X and Y pre-rotations. Owing to the time needed to perform these measurements (about 2 h), the frequency of the qubits was calibrated after every 100 repetitions. The real components of the reconstructed density matrices of the four Bell states ($(|00\rangle \pm |11\rangle)/\sqrt{2}$, $(|01\rangle \pm |10\rangle)/\sqrt{2}$) are shown in Fig. 2b–e. The state fidelities $F = \langle \psi | \rho | \psi \rangle$ between the measured density matrix (ρ) and the target Bell state (ψ) range from 85% to 89% and the concurrences from 73% to 82%, demonstrating entanglement. A parallel experiment reported a Bell-state fidelity of 78%²⁹.

To test the programmability of the two-qubit quantum processor we perform the Deutsch–Jozsa³⁰ and Grover search³¹ quantum algorithms.

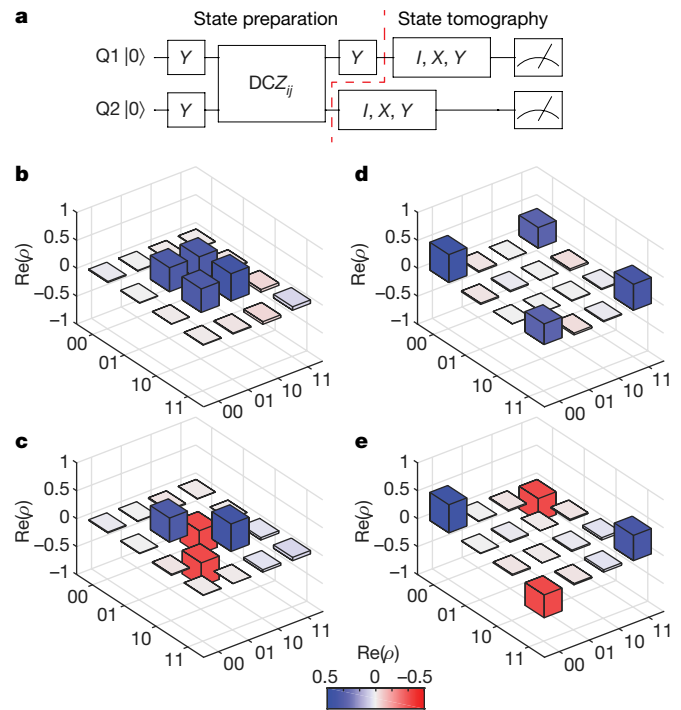


Figure 2 | Preparation of the Bell states and two-qubit entanglement in silicon. **a**, The quantum circuit used to prepare the Bell states and to perform quantum state tomography. The Bell states are prepared using a combination of single-qubit Y gates and two-qubit DCZ_{ij} gates while quantum-state tomography is performed by using different combinations of I , X and Y pre-rotation gates, where I corresponds to the identity. **b–e**, The real component of the reconstructed density matrices ($\text{Re}(\rho)$), determined using a maximum-likelihood estimate for the four Bell states $\Psi^+ = (|01\rangle + |10\rangle)/\sqrt{2}$ (**b**), $\Psi^- = (|01\rangle - |10\rangle)/\sqrt{2}$ (**c**), $\Phi^+ = (|00\rangle + |11\rangle)/\sqrt{2}$ (**d**), $\Phi^- = (|00\rangle - |11\rangle)/\sqrt{2}$ (**e**). The imaginary components of the elements of the density matrices are less than 0.08 in all cases (Supplementary Information, section S3). We measure state fidelities of $F_{\Psi^+} = 0.88 \pm 0.02$, $F_{\Psi^-} = 0.88 \pm 0.02$, $F_{\Phi^+} = 0.85 \pm 0.02$ and $F_{\Phi^-} = 0.89 \pm 0.02$, and concurrences of $c_{\Psi^+} = 0.80 \pm 0.03$, $c_{\Psi^-} = 0.82 \pm 0.03$, $c_{\Phi^+} = 0.73 \pm 0.03$ and $c_{\Phi^-} = 0.79 \pm 0.03$; all errors are 1σ from the mean.

The Deutsch–Jozsa algorithm determines whether a function is constant ($f_1(0) = f_1(1) = 0$ or $f_2(0) = f_2(1) = 1$) or balanced ($f_3(0) = 0$, $f_3(1) = 1$ or $f_4(0) = 1$, $f_4(1) = 0$). These four functions are mapped onto the unitary operators $U_{f_1} = I$, $U_{f_2} = X_2^2$, $U_{f_3} = \text{CNOT} = Y_2\text{CZ}_{11}\bar{Y}_2$ and $U_{f_4} = \text{Z-CNOT} = \bar{Y}_2\text{CZ}_{00}Y_2$, where the overbar denotes a negative rotation. For both the controlled-NOT (CNOT) and the zero-controlled-NOT (Z-CNOT) gates, the target qubit is Q2. At the end of the sequence the input qubit (Q1) will be in the $|0\rangle$ or $|1\rangle$ state for the constant and balanced functions, respectively. Grover’s search algorithm provides an optimal method for finding the unique input value x_0 of a function $f(x)$ that gives $f(x_0) = 1$ and $f(x) = 0$ for all other values of x . In the two-qubit version of this algorithm there are four input values, $x \in \{00, 01, 10, 11\}$, resulting in four possible functions $f_{ij}(x)$, with $i, j \in \{0, 1\}$. These functions are mapped onto the unitary operators, $\text{CZ}_{ij}|x\rangle = (-1)^{f_{ij}(x)}|x\rangle$, which mark the input state with a negative phase if $f_{ij}(x) = 1$. The algorithm finds the state that has been marked and outputs it at the end of the sequence.

In Fig. 3 we show the measured two-spin probabilities as a function of time during the algorithms for each function. The experimental results (circles) are in good agreement with the simulated ideal cases (dashed lines). Although a number of repetitions are needed as a result of gate and readout errors, the algorithms are successful at determining the balanced and constant functions and finding the marked state in the oracle functions. The data shown as square symbols are taken shortly after calibration and are in line with the experimental results (circles),

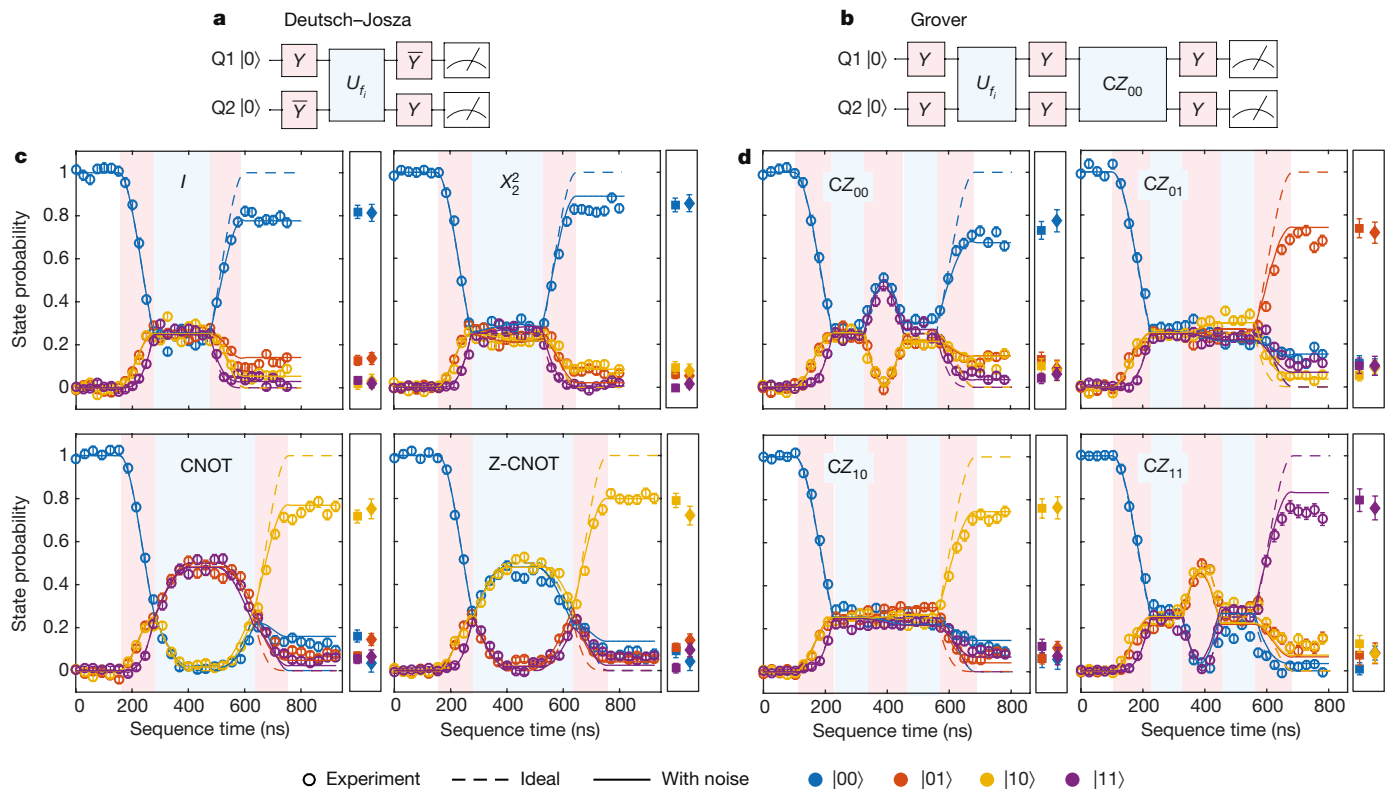


Figure 3 | Two-qubit quantum algorithms in silicon. **a, b**, Quantum circuits showing the sequences of single- and two-qubit gates that were applied for the Deutsch-Josza algorithm (**a**) and the Grover search algorithm (**b**) for two qubits. The Deutsch-Josza algorithm determines whether the function U_f is constant or balanced; the Grover search algorithm finds the state marked by the function U_f . **c, d**, Two-spin probabilities as a function of time throughout the sequence during the Deutsch-Josza algorithm (**c**) and the Grover search algorithm (**d**), for each of four possible functions. Each experimental data point (circles) corresponds to 4,000 repetitions and has been normalized to remove readout errors. The dashed lines are the simulated ideal cases and the solid

lines are the simulated results when decoherence is introduced by including quasi-static nuclear-spin and charge noise ($\sigma_e = 11 \mu\text{eV}$). For both algorithms, the square data points show the final results of the algorithms when all four functions are evaluated in the same measurement run with identical calibration. The diamonds show the result of both algorithms when using the DCZ gate, demonstrating similar performance to the CZ gate. For the Deutsch-Josza algorithm, the identity I is implemented as either a 200-ns wait (circles and squares) or as $I = X_1^4 X_2^4$ (diamonds). All error bars are 1σ from the mean. The red and blue shading in **c** and **d** corresponds to that in **a** and **b**.

indicating that calibrations remain stable throughout the hour of data collection. The diamonds show the outcome of the algorithms using the DCZ gate. In most cases, the diamonds also show similar values to the circles, which means that the DCZ gate does not improve the final result. This suggests that low-frequency single-qubit noise during the CZ gate is not dominant. The substantial difference between Hahn-echo and Ramsey decay times still points to substantial low-frequency noise. Single-qubit low-frequency noise, whether from nuclear-spin or charge noise, reduces single-qubit coherence, in particular during wait and idle times in the algorithms. In addition, charge noise affects the coupling strength J during the implementation of the CZ gates. Numerical simulations (solid lines in Fig. 3c, d and Extended Data Fig. 10) show that quasi-static nuclear-spin and charge noise can reproduce most of the features seen in the data for the two-qubit algorithm (Methods). Smaller error contributions include residual coupling during single-qubit operations and miscalibrations.

Substantial improvements could be made in the performance of the processor by using isotopically purified ^{28}Si (refs 18, 19, 21), which would increase the coherence times of the qubits. Furthermore, recent experiments have shown that symmetrically operating an exchange gate by pulsing the tunnel coupling rather than the detuning leads to a gate that is less sensitive to charge noise, greatly improving fidelities^{32,33}. With these improvements, and combined with more reproducible and scalable device structures, quantum computers with multiple qubits and fidelities above the fault-tolerance threshold should be realizable.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August 2017; accepted 16 January 2018.

Published online 14 February 2018.

- Barends, R. *et al.* Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
- Debnath, S. *et al.* Demonstration of a small programmable quantum computer with atomic qubits. *Nature* **536**, 63–66 (2016).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Maurand, R. *et al.* A CMOS silicon spin qubit. *Nat. Commun.* **7**, 13575 (2016).
- Vandersypen, L. M. K. *et al.* Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *npj Quantum Inf.* **3**, 34 (2017).
- Shulman, M. D. *et al.* Demonstration of entanglement of electrostatically coupled singlet-triplet qubits. *Science* **336**, 202–205 (2012).
- Kim, D. *et al.* Quantum control and process tomography of a semiconductor quantum dot hybrid qubit. *Nature* **511**, 70–74 (2014).
- Veldhorst, M. *et al.* A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015).
- Medford, J. *et al.* Self-consistent measurement and state tomography of an exchange-only spin qubit. *Nat. Nanotechnol.* **8**, 654–659 (2013).
- Vandersypen, L. M. K. *et al.* Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
- DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Gulde, S. *et al.* Implementation of the Deutsch-Josza algorithm on an ion-trap quantum computer. *Nature* **421**, 48–50 (2003).
- van der Sar, T. *et al.* Decoherence-protected quantum gates for a hybrid solid-state spin register. *Nature* **484**, 82–86 (2012).

14. Bluhm, H. *et al.* Dephasing time of GaAs electron-spin qubits coupled to a nuclear bath exceeding 200 μ s. *Nat. Phys.* **7**, 109–113 (2011).
15. Petta, J. R. *et al.* Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
16. Zwanenburg, F. A. *et al.* Silicon quantum electronics. *Rev. Mod. Phys.* **85**, 961–1019 (2013).
17. Tyryshkin, A. M. *et al.* Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2012).
18. Veldhorst, M. *et al.* An addressable quantum dot qubit with fault-tolerant control-fidelity. *Nat. Nanotechnol.* **9**, 981–985 (2014).
19. Muhonen, J. T. *et al.* Storing quantum information for 30 seconds in a nanoelectronic device. *Nat. Nanotechnol.* **9**, 986–991 (2014).
20. Kawakami, E. *et al.* Gate fidelity and coherence of an electron spin in an Si/SiGe quantum dot with micromagnet. *Proc. Natl Acad. Sci. USA* **113**, 11738–11743 (2016).
21. Yoneda, J. *et al.* A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102–106 (2018).
22. Kawakami, E. *et al.* Electrical control of a long-lived spin qubit in a Si/SiGe quantum dot. *Nat. Nanotechnol.* **9**, 666–670 (2014).
23. Elzerman, J. M. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).
24. Srinivasa, V., Nowack, K. C., Shafiei, M., Vandersypen, L. M. K. & Taylor, J. M. Simultaneous spin-charge relaxation in double quantum dots. *Phys. Rev. Lett.* **110**, 196803 (2013).
25. Pioro-Ladrière, M. *et al.* Electrically driven single-electron spin resonance in a slanting Zeeman field. *Nat. Phys.* **4**, 776–779 (2008).
26. Knill, E. *et al.* Randomized benchmarking of quantum gates. *Phys. Rev. A* **77**, 012307 (2008).
27. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
28. Meunier, T., Calado, V. E. & Vandersypen, L. M. K. Efficient controlled-phase gate for single-spin qubits in quantum dots. *Phys. Rev. B* **83**, 121403 (2011).
29. Zajac, D. M. *et al.* Resonantly driven CNOT gate for electron spins. *Science* **359**, 439–442 (2018).
30. Deutsch, D. & Jozsa, R. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* **439**, 553–558 (1992).
31. Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
32. Reed, M. D. *et al.* Reduced sensitivity to charge noise in semiconductor spin qubits via symmetric operation. *Phys. Rev. Lett.* **116**, 110402 (2016).
33. Martins, F. *et al.* Noise suppression using symmetric exchange gates in spin qubits. *Phys. Rev. Lett.* **116**, 116801 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was sponsored by the Army Research Office (ARO) under grant numbers W911NF-17-1-0274 and W911NF-12-1-0607. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ARO or the US Government. The US Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein. Development and maintenance of the growth facilities used for fabricating samples is supported by DOE (DE-FG02-03ER46028). We acknowledge the use of facilities supported by NSF through the University of Wisconsin-Madison MRSEC (DMR-1121288). E.K. was supported by a fellowship from the Nakajima Foundation. We acknowledge financial support from the Marie Skłodowska-Curie actions—Nanoscale solid-state spin systems in emerging quantum technologies—Spin-NANO, grant agreement number 676108. We acknowledge discussions with S. Dobrovitski, C. Dickel, A. Rol, J. P. Dehollain, Z. Ramlakhan and members of the Vandersypen group, and technical assistance from R. Schouten, R. Vermeulen, M. Tiggelman, M. Ammerlaan, J. Haanstra, R. Roeleveld and O. Benningshof.

Author Contributions T.F.W. performed the experiment with help from E.K. and P.S., T.F.W. and S.G.J.P. analysed the data, S.G.J.P. performed the simulations of the algorithms, T.F.W., S.G.J.P., E.K., P.S., M.V., M.F., S.N.C., M.A.E. and L.M.K.V. contributed to the interpretation of the data and commented on the manuscript, D.R.W. fabricated the device, D.E.S. and M.G.L. grew the Si/SiGe heterostructure, T.F.W. wrote the manuscript (S.G.J.P. wrote parts of Methods) and L.M.K.V. conceived and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.M.K.V. (l.m.k.vandersypen@tudelft.nl).

Reviewer Information Nature thanks H. Bluhm and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Estimate of initialization and readout errors for Q1 and Q2. The initialization and readout procedures for Q1 and Q2 are described in Extended Data Fig. 2. The initialization and readout fidelities of Q2 were extracted by performing the following three experiments and measuring the resulting spin-up probabilities (P_1 , P_2 and P_3): (i) initialize Q2 and wait for $7T_1$; (ii) initialize Q2; and (iii) initialize and perform a π rotation on Q2. These three spin-up probabilities are related to the initialization fidelity (γ_2) and the spin-up and spin-down readout fidelities ($F_{|0\rangle,2}$ and $F_{|1\rangle,2}$) by

$$\begin{aligned} P_1 &= 1 - F_{|0\rangle,2} \\ P_2 &= F_{|1\rangle,2}(1 - \gamma_2) + (1 - F_{|0\rangle,2})\gamma_2 \\ \frac{P_3}{P_{\pi 2}} &= F_{|1\rangle,2}\gamma_2 + (1 - F_{|0\rangle,2})(1 - \gamma_2) \end{aligned} \quad (2)$$

where $P_{\pi 2}$ is the expected probability to be in the spin-up state after the application of the π pulse for Q2, which is determined as described below. In equation (2) we assume that a waiting time of $7T_1$ leads to 100% initialization and that the measured spin-up counts are due to the readout infidelity. By solving these equations we can extract the initialization and readout fidelities. For Q1, we performed initialization by pulsing to a spin-relaxation hotspot (Extended Data Fig. 5) for $500T_1$ and we therefore assume that the initialization fidelity is about 100%. Consequently, the readout fidelities of Q1 were extracted by performing only experiments (ii) and (iii) above. The readout and initialization fidelities for Q1 (Q2) during the state tomography experiments were estimated to be $\gamma_1 > 99\%$ ($\gamma_2 > 99\%$), $F_{|0\rangle,1} = 92\%$ ($F_{|0\rangle,2} = 86\%$) and $F_{|1\rangle,1} = 54\%$ ($F_{|1\rangle,2} = 76\%$), for which we used $P_{\pi 1} = 98\%$ ($P_{\pi 2} = 97\%$) on the basis of simulations that include the dephasing time of the qubits (see below). The average measurement fidelity $F_m = (F_{|0\rangle} + F_{|1\rangle})/2$ for Q1 (Q2) is 73% (81%). These fidelities are limited mostly by the finite electron temperature $T_e \approx 130$ mK and the fast spin-relaxation time of Q2 ($T_1 = 3.7$ ms), which is probably caused by a spin-relaxation hotspot that is due to a similar valley splitting and Zeeman energy³⁶.

Removing readout errors from the measured two-spin probabilities. In the experiment the measured two-spin probabilities $\mathbf{P}^M = (P_{|00\rangle}^M, P_{|01\rangle}^M, P_{|10\rangle}^M, P_{|11\rangle}^M)^T$ include errors due to the limited readout fidelities $F_{|0\rangle,i}$ and $F_{|1\rangle,i}$ of spin-down (0) and spin-up (1) electrons for qubit i . To remove these readout errors and obtain the actual two-spin probabilities $\mathbf{P} = (P_{|00\rangle}, P_{|01\rangle}, P_{|10\rangle}, P_{|11\rangle})^T$, we use the relationship $\mathbf{P}^M = (\hat{F}_1 \otimes \hat{F}_2)\mathbf{P}$, where

$$\hat{F}_i = \begin{pmatrix} F_{|0\rangle,i} & 1 - F_{|1\rangle,i} \\ 1 - F_{|0\rangle,i} & F_{|1\rangle,i} \end{pmatrix} \quad (3)$$

State tomography. The density matrix of a two-qubit state can be expressed as $\rho = \sum_{i=1}^{16} c_i M_i$, where M_i are 16 linearly independent measurement operators. The coefficients c_i are calculated from the expectation values m_i of the measurement operators using a maximum-likelihood estimate^{11,37}. The expectation values were calculated by performing 16 combinations of (I , X , Y , X^2) pre-rotations on Q1 and Q2 and measuring the two-spin probabilities over 10,000 repetitions per measurement. The two-spin probabilities were converted to actual two-spin probabilities by removing the readout errors using equation (3). To calculate the density matrices shown in Fig. 2 we used only the data from the (I , X , Y) pre-rotations on the assumption that I will give a more accurate estimate of the expectation values than X^2 owing to gate infidelities. If we include X^2 then we achieve state fidelities of 80%–84% and concurrences of 67%–71% (Supplementary Information, section S3). In the analysis we assume that the pre-rotations are perfect, which is a reasonable approximation owing to the high single-qubit Clifford-gate fidelities (98%) compared to the measured state fidelities (85%–89%). The state tomography experiment was performed in parallel with the fidelity experiments described above and a Ramsey experiment that was used to calibrate the frequency actively.

Error analysis. Error analysis was performed using a Monte Carlo method by assuming a multinomial distribution for the measured two-spin probabilities and a binomial distribution for the probabilities P_1 , P_2 and P_3 used to calculate the fidelities. Values from these distributions were sampled randomly and the procedures outlined above were followed. This was repeated 250 times to build up the final distributions that we used to determine the mean values and the standard deviation.

Simulation of two electron spins in a double quantum dot. In the simulation, we consider two electrons in two tunnel-coupled quantum dots, with an external magnetic field B_0 applied to both dots. In addition to this field, the two dots have different Zeeman energies owing to the magnetic-field gradient across the double quantum dot that is generated by the micromagnets. The Zeeman energy of Q1 (Q2) is denoted as E_{Z1} (E_{Z2}). The double-quantum-dot system is modelled by the Hamiltonian³⁸

$$\hat{H} = \begin{pmatrix} -\beta & 0 & 0 & 0 & 0 & 0 \\ 0 & -\Delta\nu & 0 & 0 & t & t \\ 0 & 0 & \Delta\nu & 0 & -t & -t \\ 0 & 0 & 0 & \beta & 0 & 0 \\ 0 & t & -t & 0 & U_1 + \varepsilon & 0 \\ 0 & t & -t & 0 & 0 & U_2 - \varepsilon \end{pmatrix}$$

with the states $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$, $S(2, 0)$ and $S(0, 2)$ as the eigenbasis. In this Hamiltonian, $\beta = (E_{Z1} + E_{Z2})/2$, $\Delta\nu = (E_{Z1} - E_{Z2})/2$, $\sqrt{2}t$ is the tunnel coupling between the $(1, 1)$ and $(0, 2)/(2, 0)$ singlet states, and U_i is the on-site charging energy of the i th quantum dot. To study the phases of the qubits during the control pulses, the Hamiltonian is transformed into a rotating frame using

$$\tilde{H} = V H V^\dagger + i\hbar(\partial_t V) V^\dagger \quad (4)$$

where $V = \exp\{-i[E_{Z1}(\hat{\sigma}_z \otimes \hat{I}) + E_{Z2}(\hat{I} \otimes \hat{\sigma}_z)]t\}$ is the matrix that describes the unitary transformation normalized such that $\hbar = 1$. The transformed Hamiltonian is

$$\tilde{H} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & te^{i\Delta\nu t} & te^{i\Delta\nu t} \\ 0 & 0 & 0 & 0 & -te^{-i\Delta\nu t} & -te^{-i\Delta\nu t} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & te^{-i\Delta\nu t} & -te^{i\Delta\nu t} & 0 & U_1 + \varepsilon & 0 \\ 0 & te^{-i\Delta\nu t} & -te^{i\Delta\nu t} & 0 & 0 & U_2 - \varepsilon \end{pmatrix}$$

To model the single-qubit gates during EDSR, we use the Hamiltonian

$$\hat{H}_{MW} = \sum_k B_{MW,k} \cos(\omega_k t + \phi_k)(\hat{\sigma}_x \otimes \hat{I} + \hat{I} \otimes \hat{\sigma}_x)$$

which assumes the same drive amplitude on each of the qubits. Here, k represents the k th signal, which has an angular frequency ω_k , phase ϕ_k and driving amplitude $B_{MW,k}$. This Hamiltonian is transformed into the rotating frame using equation (4), and the rotating-wave approximation can be made to remove the fast-driving elements because the Rabi frequency is much smaller than the Larmor precession. Doing so gives the Hamiltonian

$$\tilde{H}_{MW} = \sum_k \begin{pmatrix} 0 & \Omega_k e^{i\Delta\omega_1 t} & \Omega_k e^{i\Delta\omega_2 t} & 0 & 0 & 0 \\ \Omega_k^* e^{-i\Delta\omega_1 t} & 0 & 0 & \Omega_k e^{i\Delta\omega_2 t} & 0 & 0 \\ \Omega_k^* e^{-i\Delta\omega_2 t} & 0 & 0 & \Omega_k e^{i\Delta\omega_1 t} & 0 & 0 \\ 0 & \Omega_k^* e^{-i\Delta\omega_2 t} & \Omega_k^* e^{-i\Delta\omega_1 t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\Omega_k = B_{MW,k} e^{i\phi_k}$, Ω_k^* is the complex conjugate of Ω_k and $\Delta\omega_k = \omega_k - \omega_{\text{qubit},i}$.

The dynamics of the two-qubit system can be described by the Schrödinger–von Neumann equation, $\rho_{t+\Delta t} = e^{-i\hat{H}\Delta t/\hbar} \rho_t e^{i\hat{H}\Delta t/\hbar}$, which we solved numerically using the Armadillo linear algebra library in C++, with the matrix exponentials solved using scaling methods ($e^A = \prod_k e^{A/2^k}$) and a Taylor expansion. In the experiments, we apply microwave pulses with square envelopes that have a finite rise time owing to the limited bandwidth of the in-phase (I) and quadrature (Q) modulation channels of the microwave vector source. For simplicity, we approximate these microwave pulses with a perfect square envelope. The detuning pulses were modelled with a finite rise and fall time using a Fermi–Dirac function to take (a) diabatic effects into account. The finite rise time was set to 2 ns on the basis of the cut-off frequency of the low-pass filter that was attached to the lines and used to pulse the detuning pulses.

Modelling noise in the simulation. In the model we include three different noise sources. The first two noise sources are from fluctuating nuclear spins in the natural silicon quantum well, which generate quasi-static magnetic noise that couples to the qubits via the $Z \otimes I$ and $I \otimes Z$ terms in the Hamiltonian. These fluctuations are treated as two independent noise sources because D1 and D2 are in different locations in the quantum well and sample the field from different nuclear spins. The third noise source is charge noise, which can couple to the qubits via the magnetic-field gradient from the micromagnets. We model this noise as magnetic noise on the $Z \otimes I$ and $I \otimes Z$ terms in the Hamiltonian. Charge noise also couples to the spins via the exchange coupling, which leads to noise on the $Z \otimes Z$ term in the Hamiltonian.

In our simulations, we treat these noise sources as quasi-static, whereby the noise is static within each cycle and changes only between measurement cycles. This approximation is reasonable because the noise in the system is pink, with

low frequencies in the power spectrum being more pronounced²⁰. The static noise due to each noise source was modelled by sampling a random value from a Gaussian distribution with standard deviation σ , corresponding to the contribution to dephasing of that noise process. After sampling the static noise, the time evolution of the qubits during a gate sequence was calculated. This time evolution was averaged over many repetitions to give the final result; for each repetition new values for the static noise were sampled. In total, for each simulation we performed 5,000 repetitions to ensure convergence.

In the experiment, single-qubit gates are performed at higher detuning near the centre of the (1, 1) charge region at a detuning of $\varepsilon = -3$ meV, where the exchange is low ($J_{\text{off}} = 270$ kHz); a two-qubit CZ gate is performed by pulsing to low detuning $\varepsilon = -0.7$ meV, where the exchange is high ($J_{\text{on}} = 6$ MHz). To estimate the relative effect of charge noise on the $Z \otimes I$, $I \otimes Z$ and $Z \otimes Z$ terms at these two detuning points, we use the spectroscopy data for the qubits as a function of detuning energy shown in Extended Data Fig. 8. The four resonances observed correspond to the four transitions shown in Extended Data Fig. 8c between the $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$ eigenstates. From fits of this data we estimate the derivative of the transition energy from state $|i\rangle$ to $|j\rangle$ at a particular detuning, $dE_{|i\rangle \rightarrow |j\rangle}/d\varepsilon|_{\varepsilon}$, which is directly proportional to the magnitude of fluctuations in the transition energy under the influence of charge noise. Fixing the energy of the $|00\rangle$ state, from these derivatives we calculate the relative noise levels on the other energy eigenstates:

$$B(\varepsilon) = \begin{pmatrix} 0 \\ \frac{\partial E_{|00\rangle \leftrightarrow |01\rangle}}{\partial \varepsilon} \Big|_{\varepsilon} \\ \frac{\partial E_{|00\rangle \leftrightarrow |10\rangle}}{\partial \varepsilon} \Big|_{\varepsilon} \\ \frac{\partial E_{|00\rangle \leftrightarrow |01\rangle}}{\partial \varepsilon} \Big|_{\varepsilon} + \frac{\partial E_{|01\rangle \leftrightarrow |11\rangle}}{\partial \varepsilon} \Big|_{\varepsilon} \end{pmatrix} \quad (5)$$

In the regime in which $J \ll \Delta v$, the Hamiltonian of the system can be approximated as $H = -E_{Z1}(Z \otimes I) - E_{Z2}(I \otimes Z) + J(Z \otimes Z) - J/4(I \otimes I)$. The relative noise on E_{Z1} , E_{Z2} and J is found by decomposing the four noise levels in equation (5) in terms of the basis $(-Z \otimes I, -I \otimes Z, Z \otimes Z, -I \otimes I/4)$ by calculating $A^{-1}B(\varepsilon)$, where

$$A = \begin{pmatrix} -1/2 & -1/2 & 1/4 & -1/4 \\ -1/2 & 1/2 & -1/4 & -1/4 \\ 1/2 & -1/2 & -1/4 & -1/4 \\ 1/2 & 1/2 & 1/4 & -1/4 \end{pmatrix}$$

We estimate the relative composition of the noise for (E_{Z1}, E_{Z2}, J) at $\varepsilon = -3$ meV to be (0.12, 0.24, 0) and at $\varepsilon = -0.7$ meV ($J = 6$ MHz) to be (0.61, 0.23, 0.26). Note that this is a crude approximation because we take into account only voltage noise along the detuning axis, whereas in reality charge noise acts also along other axes. We do not include calibration errors in the simulation. On the basis of the AllXY and Ramsey calibration experiments (Supplementary Information, section S2), miscalibrations of a few per cent are possible.

Estimating charge noise from the decay of the decoupled CZ oscillations. The dephasing due to charge noise coupling into the double-quantum-dot system via the exchange energy is measured by varying the duration of the DCZ gate between two $\pi/2$ pulses on Q1 (Extended Data Fig. 9) for $J = 6$ MHz. The DCZ gate removes the effect of quasi-static noise on the $Z \otimes I$ and $I \otimes Z$ terms in the Hamiltonian, and the decay time of the oscillations $T_2 = 1,640$ ns is assumed to be due to noise on the $Z \otimes Z$ term. The data are fitted using either a Gaussian (black line) or exponential decay (red line). The exponential decay seems to fit better to the data, which suggests that either higher-frequency noise plays a part³⁹ or the origin of the noise is from a few two-level fluctuators⁴⁰. Because the decay time of the DCZ gate is longer than that of the CZ gate, there is also a substantial quasi-static noise contribution.

For simplicity, we include only the quasi-static contribution in our noise model. For Gaussian quasi-static noise with standard deviation σ_{ε} , the decay time is

$$1/T_2 = \frac{1}{2} \frac{\partial J}{\partial \varepsilon} \Big|_{\varepsilon} \frac{\sigma_{\varepsilon}}{\sqrt{2}\hbar}$$

The factor of 1/2 is needed because it is the noise on $J/2$ that contributes to the decay. This is because the target qubit precesses with a frequency of $J/2$ (ignoring the $I \otimes Z$ and $Z \otimes I$ terms) when the control qubit is in an eigenstate. From the dephasing time and $dJ/d\varepsilon|_{\varepsilon} = 1.0 \times 10^{-4}$ extracted from Extended Data Fig. 8a, b, we estimate the charge noise on the detuning to be 11 μ eV. The data in Extended Data Fig. 9 used to extract this value were collected over about 40 min with no active calibration on the detuning pulse. The time needed for each single-shot measurement was around 10 ms.

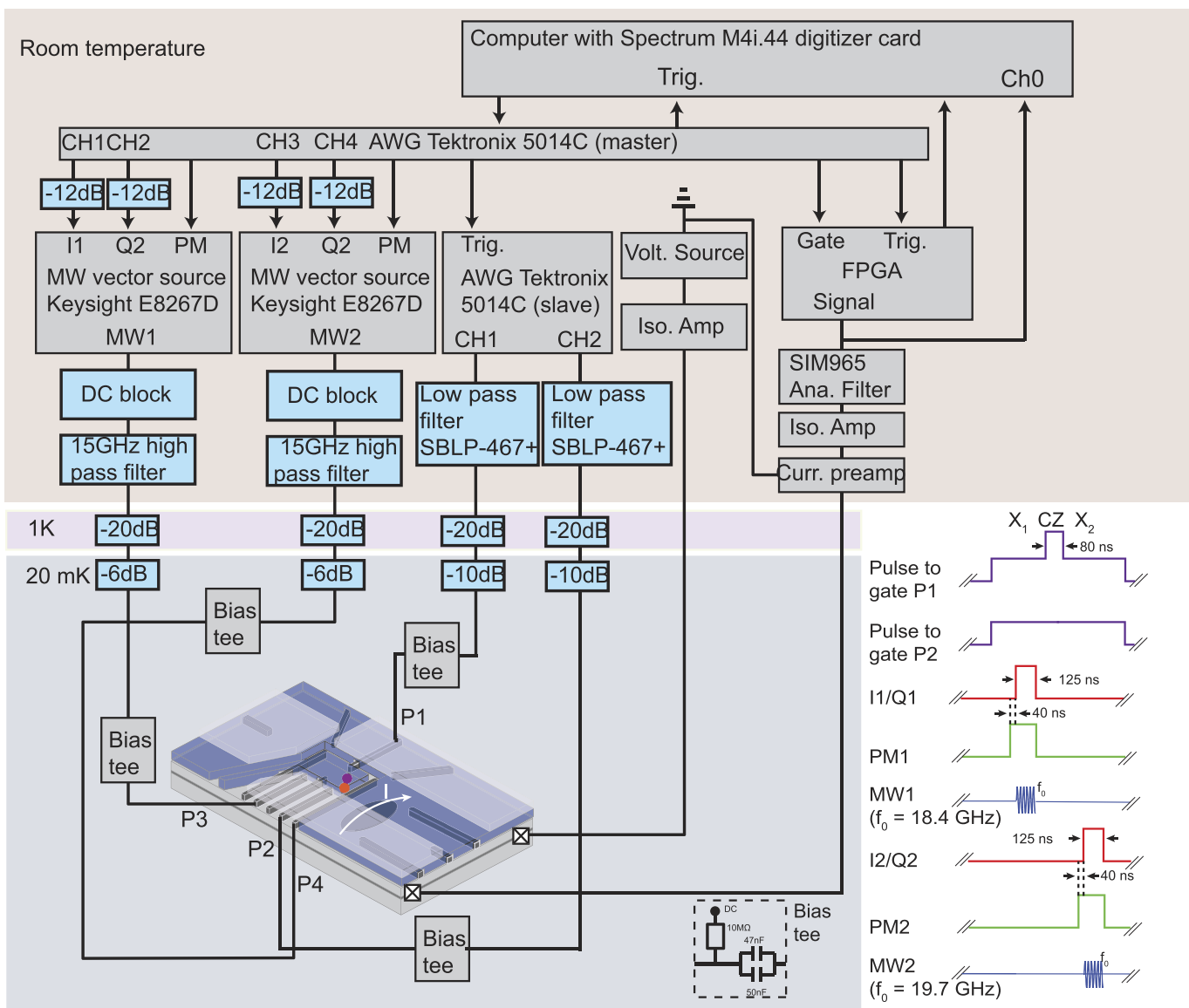
Simulations of the two-qubit algorithms. To describe the double-quantum-dot system used in the experiment, we use the following parameters in the Hamiltonian. The qubit frequencies were chosen to be $E_{Z1}/\hbar = 18.4$ GHz and $E_{Z2}/\hbar = 19.7$ GHz, and the on-site charging energies to be $U_1 = U_2 = 3.5$ meV, comparable to the experimental values. The tunnel coupling was chosen to be $t = 210$ MHz so that the residual exchange energy J_{off} was 300 kHz, similar to that measured in the experiment. The two-qubit gates were implemented by choosing a value of ε for which $J = 6$ MHz when diagonalizing the Hamiltonian \hat{H} .

The results of the simulations for the Deutsch-Jozsa algorithm and the Grover algorithm using the CZ gate and the DCZ gate are shown in Fig. 3 and Extended Data Fig. 10. The amplitudes for the three noise sources used in the simulations were identical for all 16 panels. The value of charge noise used was 11 μ eV (see above), whereas the nuclear-spin noise for Q1 and Q2 was chosen to give the single-qubit decoherence times $T_2^* = 1,000$ ns and $T_2^* = 600$ ns measured in the Ramsey experiment (Extended Data Fig. 3). This gave a dephasing time of Q1 (Q2) due to nuclear spin of $T_{2\text{nuc}}^* = 1,200$ ns ($T_{2\text{nuc}}^* = 800$ ns). The simulations reproduce many of the features of the experimental data for the algorithms.

By simulating the algorithms, we learn that the residual exchange coupling J_{off} during single-qubit gates has little effect (less than 2%) on the result of the algorithms. Furthermore, we find that without noise on the single-qubit terms it is difficult to get consistent agreement with the data. Additional noise on the coupling strength improves the agreement. In contrast to the cases of the Deutsch-Jozsa algorithm and the conventional Grover algorithm, the simulation for the decoupled version of Grover's algorithm predicts a better outcome than the experiment. This case uses the longest sequence of operations, leaving the most room for discrepancies between model and experiment to build up. These discrepancies could have several origins: (i) the implementation of the static noise model not being accurate enough; (ii) non-static noise having a role; (iii) the calibration errors in the gates that were left out of the simulation; or (iv) variations in the qubit parameters and noise levels between experiments. Finally, we note that initialization and readout errors are not taken into account in the simulations. Because initialization errors are negligible and the data shown were renormalized to remove the effect of readout errors, the simulated and experimental results can be compared directly.

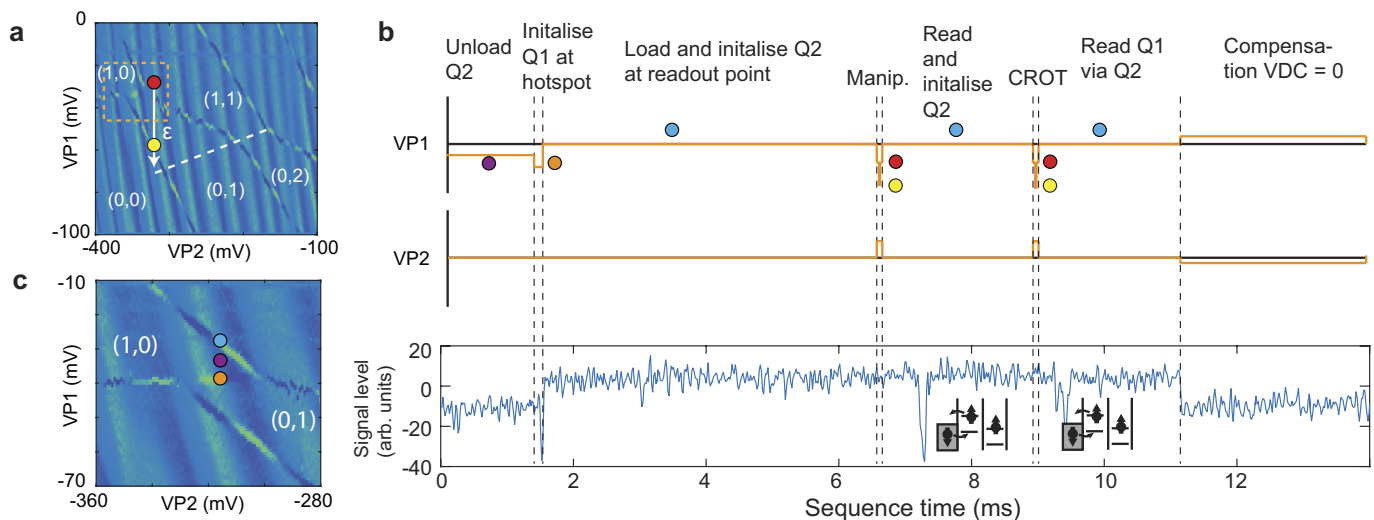
Data availability. The raw data and analysis that support the findings of this study are available in the Zenodo repository (<https://doi.org/10.5281/zenodo.1135014>).

36. Yang, C. H. *et al.* Spin-valley lifetimes in a silicon quantum dot with tunable valley splitting. *Nat. Commun.* **4**, 2069 (2013).
37. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
38. Das Sarma, S., Wang, X. & Yang, S. Hubbard model description of silicon spin qubits: charge stability diagram and tunnel coupling in Si double quantum dots. *Phys. Rev. B* **83**, 235314 (2011).
39. Dial, O. E. *et al.* Charge noise spectroscopy using coherent exchange oscillations in a singlet-triplet qubit. *Phys. Rev. Lett.* **110**, 146804 (2013).
40. Ithier, G. *et al.* Decoherence in a superconducting quantum bit circuit. *Phys. Rev. B* **72**, 134519 (2005).



Extended Data Figure 1 | Schematic of the measurement set-up. The sample was bonded to a printed circuit board (PCB) mounted onto the mixing chamber of a dilution refrigerator. All measurements were performed at the base temperature of the fridge, $T_{\text{base}} \approx 20$ mK. DC voltages were applied to all of the gate electrodes using room-temperature digital-to-analogue converters via filtered lines (not shown). Voltage pulses were applied to plunger gates P1 and P2 using a Tektronix 5014C arbitrary waveform generator (AWG) with 1-GHz clock rate. The signals from the AWGs passed through a room-temperature low-pass filter and attenuators at different stages of the fridge and were added to the DC signals via bias tees mounted on the PCB. Two Keysight E8267D vector microwave sources, MW1 and MW2, were used to apply microwaves (18–20 GHz) to perform EDSR on Q1 and Q2, respectively. The signals passed through room-temperature DC blocks and custom-built 15-GHz high-pass filters and attenuators at different stages of the fridge and were added to the DC signals via bias tees mounted on the PCB. The output of the microwave source (phase, frequency, amplitude and duration) was controlled with I/Q vector modulation. The I/Q signals were generated with another Tektronix 5041C, which was the master device for the entire set-up and provided trigger signals for the other devices. In addition to the

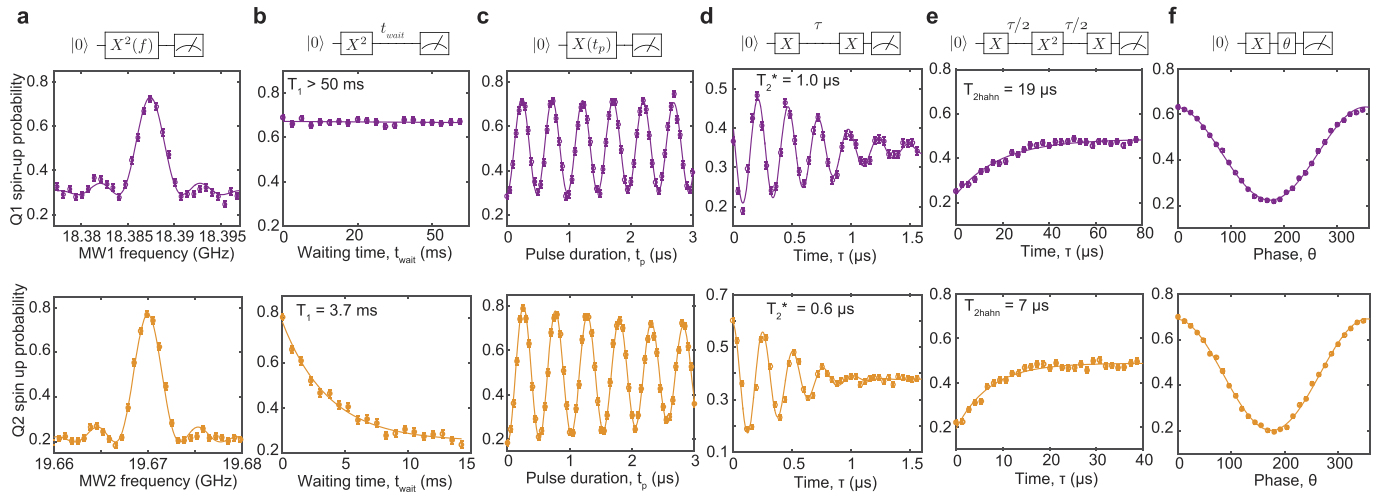
vector modulation, we used pulse modulation to give an on/off microwave power output ratio of 120 dB. Although I/Q modulation can be used to output multiple frequencies, the bandwidth of the AWG was not sufficient to control both qubits with one microwave source owing to their large separation in frequency (1.3 GHz). The sensor current I was converted to a voltage signal with a custom-built preamplifier, and an isolation amplifier was used to separate the signal ground, with the measurement equipment grounded to reduce interference. Following this, a 20-kHz Bessel low-pass filter was applied to the signal using a SIM965 analogue filter. A field-programmable gate array (FPGA) analysed the voltage signal during the readout and assigned the trace to be spin-up if the voltage fell below a certain threshold. The voltage signal could also be measured with a digitizer card in the computer. The shapes of the pulses generated by the AWGs and microwave sources during qubit manipulation, along with the typical timescales, are shown in the lower right. Square pulses were used to perform the CZ gate and as the input for the I/Q modulation to generate microwave pulses. The pulse modulation was turned on 40 ns before turning on the I/Q signal, owing to the time needed for the modulation to switch on.



Extended Data Figure 2 | Measurement protocol for two electron spins.

a, Stability diagram of the double quantum dot, showing the positions in gate space used to perform single-qubit gates (red circle) and two-qubit gates (yellow circle). The map shows the numerically obtained differential current dI/dV_{P1} through the quantum-dot charge sensor as a function of the voltages on P1 and P2. The white dashed line is the (1, 1)–(0, 2) inter-dot transition line. The white arrow indicates the detuning axis ε used in the experiments. Although the detuning pulse for the two-qubit gate crosses the charge-addition lines of D1 and D2, the quantum dots remain in the (1, 1) charge state because the pulse time is much shorter than the electron tunnel times to the reservoirs. **b**, Plot of the voltage pulses applied to plunger gates P1 and P2 and the response of the quantum-dot charge sensor over one measurement cycle. D2 is unloaded by pulsing into the (1, 0) charge region for 1.5 ms (purple circle). The electron on D1 is initialized to spin-down by pulsing to a spin-relaxation hotspot at the (1, 0) and (0, 1) charge degeneracies (orange circle) for 50 μ s (see Extended Data Fig. 5). D2 is loaded with a spin-down electron by pulsing to the readout position for 4 ms (blue circle). During manipulation, the voltages

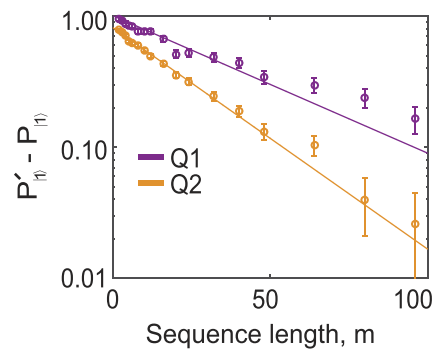
on the plunger gates are pulsed to the red circle for single-qubit gates and to the yellow circle for two-qubit gates, where the exchange is about 6 MHz. After manipulation, the spin of the electron on D2 is measured by pulsing to the readout position (blue circle) for 0.7 ms, where the Fermi level of the reservoir is between the spin-up and spin-down electrochemical potentials of D2. If the electron is spin-up then it can tunnel out, after which a spin-down electron tunnels back in. These two tunnel events are detected by the quantum-dot sensor as a single blip in the current signal. An additional 1.3 ms is spent at the readout position so that D2 is initialized to spin-down with high fidelity. Following this, Q1 is measured by first performing a CROT gate at the yellow circle so that $\alpha|00\rangle + \beta|10\rangle \xrightarrow{\text{CROT12}} \alpha|00\rangle + \beta|11\rangle$, where CROT12 indicates a CROT gate with Q1 as the control and Q2 as the target. A projective measurement of Q1 is then performed by measuring Q2 at the readout position for 0.7 ms (blue circle). Finally, we add a compensation pulse to VP1 and VP2 so that over the measurement cycle $V_{DC} = 0$ to mitigate charging effects in the bias tees. **c**, Close-up of the stability diagram in **a** showing the positions in gate space used for initialization and readout.



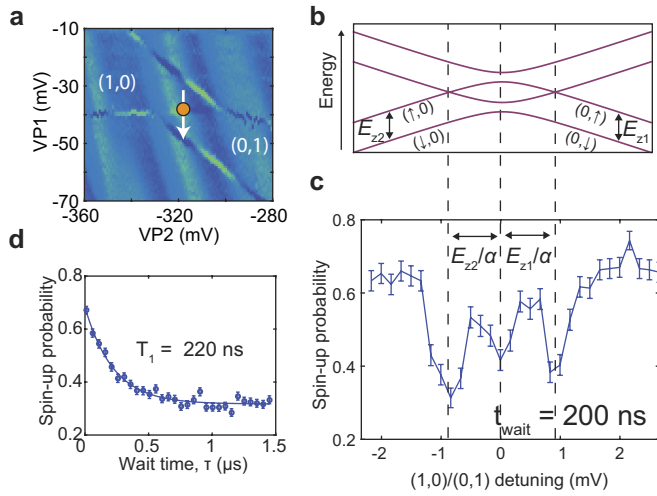
Extended Data Figure 3 | Single-qubit properties and two-axis control.

The purple (top) and orange (bottom) data points correspond to measurements performed on Q1 and Q2, respectively, in the (1, 1) regime (red circle in Extended Data Fig. 2). **a**, Spin-up fraction as a function of the microwave frequency of an applied π pulse, showing a resonant frequency of 18.387 GHz (19.670 GHz) for Q1 (Q2). **b**, The spin-relaxation time is measured by preparing the qubit to spin-up and varying the wait time before readout. From the exponential decay in the spin-up probability we measure $T_1 > 50$ ms ($T_1 = 3.7 \pm 0.5$ ms) for Q1 (Q2). **c**, Spin-up probability as a function of microwave duration, showing Rabi oscillations of 2.0 MHz for Q1 and Q2. **d**, The dephasing time is measured by applying a Ramsey pulse sequence and varying the free evolution time τ . Oscillations were added artificially to improve the fit of the decay by making the phase of the

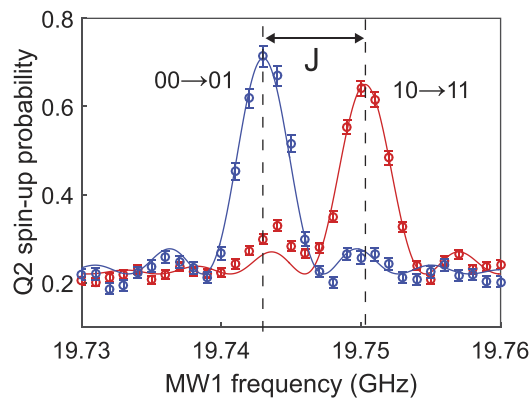
last microwave pulse dependent on the free evolution time, $\phi = \sin(\omega\tau)$, where $\omega = 4$ MHz. By fitting the data with a Gaussian decay, $P_{|1\rangle} \propto \exp[-(\tau/T_2^*)^2] \sin(\omega\tau)$, we extract $T_2^* = 1.0 \pm 0.1$ μ s ($T_2^* = 0.6 \pm 0.1$ μ s) for Q1 (Q2). In the measurement for Q1, the first $\pi/2$ microwave pulse is a Y gate. The Ramsey measurement was performed over about 20 min with the frequency calibrated every approximately 1 min. **e**, The coherence time of Q1 (Q2) can be extended to $T_{2\text{Hahn}} = 19 \pm 3$ μ s ($T_{2\text{Hahn}} = 7 \pm 1$ μ s) by a Hahn echo sequence. The coherence time is extracted from an exponential fit to the spin-up probability as a function of the free evolution time in the Hahn echo sequence. **f**, Full two-axis control is demonstrated by applying two $\pi/2$ pulses and varying the phase of the second one. All error bars are 1σ from the mean.



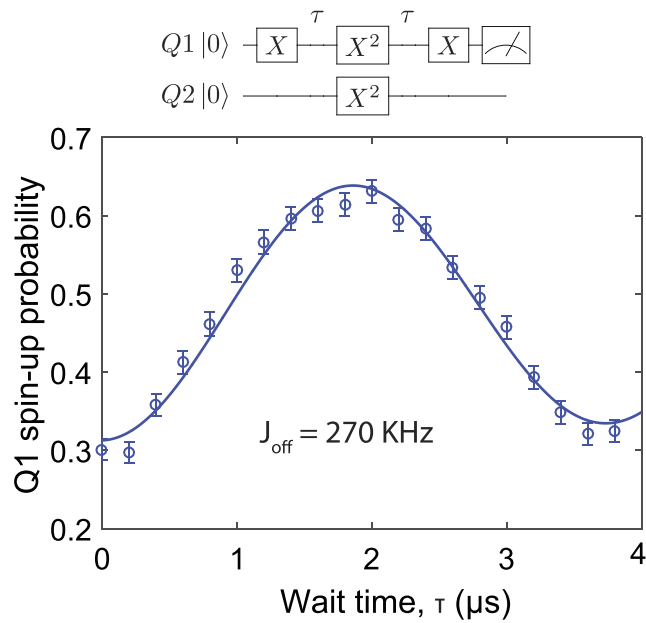
Extended Data Figure 4 | Randomized benchmarking of single-qubit gates. Randomized benchmarking of the single-qubit gates for each qubit is performed by applying a randomized sequence of a varying number of Clifford gates m to either the $|1\rangle$ or $|0\rangle$ state and measuring the final spin-up probability $P'_{|1\rangle}$ or $P_{|1\rangle}$, respectively. All gates in the Clifford group are decomposed into gates from the set $\{I, \pm X, \pm X^2, \pm Y, \pm Y^2\}$. The purple (orange) data points show the difference in the spin-up probabilities $P'_{|1\rangle} - P_{|1\rangle}$ for Q1 (Q2) as a function of sequence length. For each sequence length m we average over 32 different randomized sequences. From an exponential fit (solid lines) of the data, $P'_{|1\rangle} - P_{|1\rangle} = ap^m$, we estimate average Clifford-gate fidelities $F_C = 1 - (1 - p)/2$ of 98.8% and 98.0% for Q1 and Q2, respectively. The last three data points from both datasets were omitted from the fits because they deviate from a single exponential²⁰. All error bars are 1σ from the mean.



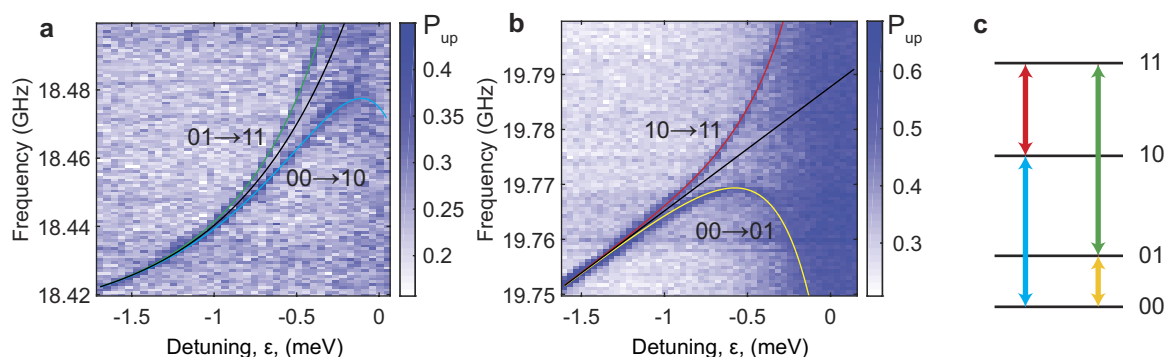
Extended Data Figure 5 | Spin-relaxation hotspots used for high-fidelity initialization. **a**, Close-up stability diagram of the (1, 0)–(0, 1) charge transition. The white arrow defines the detuning axis between D1 and D2 controlled with P1. **b**, Schematic of the energy-level diagram as a function of detuning for one electron spin in a double quantum dot. **c**, Spin-relaxation hotspots are measured by preparing the electron on D1 to spin-up using EDSR, applying a voltage pulse along the detuning axis (white arrow in **a**) for a wait time of 200 ns and performing readout of the electron spin. We observe three dips in the spin-up probability, which correspond to spin-relaxation hot spots. The first and third hotspot are due to anticrossings between the (0, ↓) and (↑, 0) states and the (↓, 0) and (0, ↑) states²⁴. The second hotspot occurs at zero detuning. The voltage separation between the first and third hot spot corresponds to the sum of the Zeeman energy of D1 and D2 divided by the gate lever arm α along the detuning axis. Knowing precisely the Zeeman energies from EDSR spectroscopy, we can accurately extract the gate lever arm to be $\alpha = 0.09e$. **d**, The spin-relaxation time at zero detuning (orange circle in **a**) is found to be $T_1 = 220$ ns by measuring the exponential decay of the spin-up probability as a function of wait time τ at zero detuning. All error bars are 1σ from the mean.



Extended Data Figure 6 | Two-qubit CROT gate. a, Microwave spectroscopy of Q2 close to zero detuning between the $(1, 1)$ and $(0, 2)$ states (yellow dot in Extended Data Fig. 2a) with the exchange coupling on. The blue and red curves show the resonance of Q2 after preparing Q1 into spin-down and spin-up, respectively. The resonance frequency of Q2 shifts by the exchange coupling, and by applying a π pulse at one of these frequencies we can perform a CROT gate, which is equivalent to a CNOT gate up to a \hat{z} rotation. As discussed in the main text, this CROT gate is used to perform the projective measurement of Q1. All error bars are 1σ from the mean.



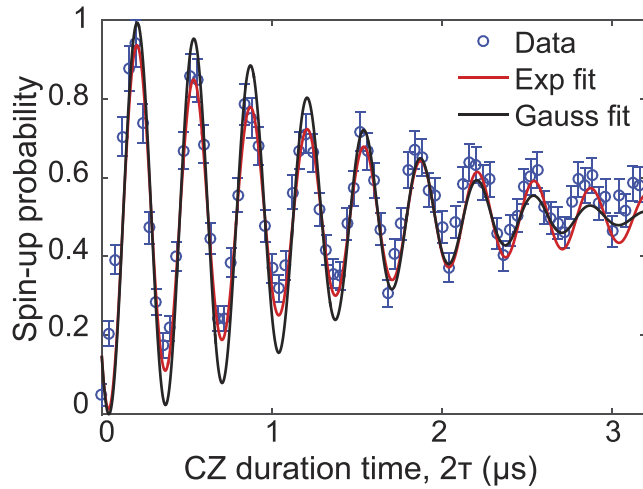
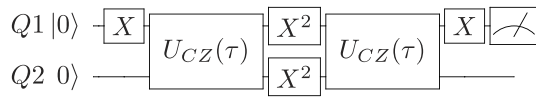
Extended Data Figure 7 | Measurement of J_{off} using a decoupling sequence. The exchange coupling J_{off} during single-qubit gates is measured using a two-qubit Hahn echo sequence, which cancels out any unconditional \hat{z} rotations during the free evolution time τ . Fitting the spin-up probability as a function of free evolution time τ using the functional form $\sin(2\pi J_{\text{off}}\tau)$, we extract $J_{\text{off}} = 270 \text{ kHz}$. All error bars are 1σ from the mean.



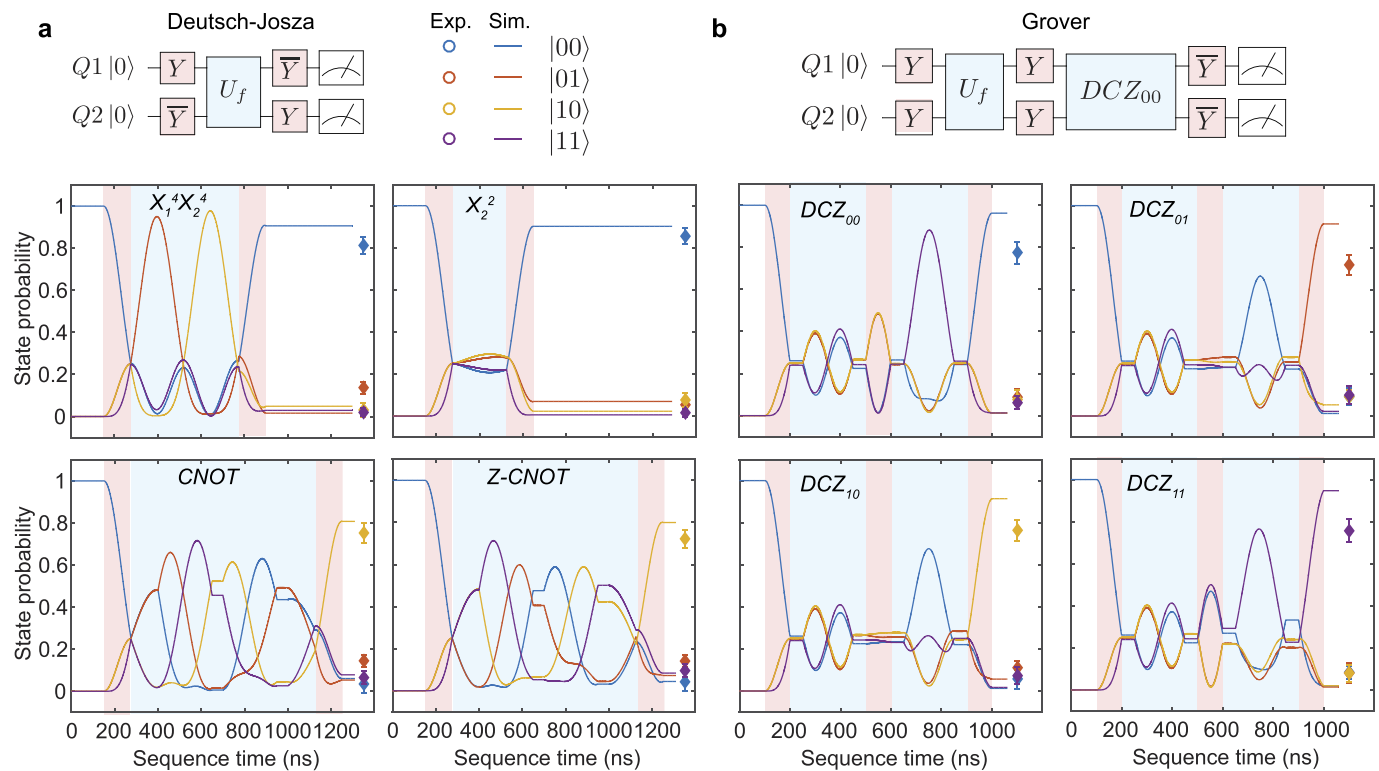
Extended Data Figure 8 | Microwave spectroscopy of Q1 and Q2.

a, b, Spectroscopy of Q1 (**a**) and Q2 (**b**) versus detuning energy ε after initializing the other qubit to $(|0\rangle + |1\rangle)/\sqrt{2}$. Towards $\varepsilon=0$ there are two resonances each for Q1 and Q2, which are separated by the exchange energy $J(\varepsilon)/h$. As discussed, the Zeeman energy $E_Z(\varepsilon)$ of Q1 and Q2 also depends on detuning because changes to the applied voltages shift the position of the electron in the magnetic-field gradient. The four resonance frequencies are fitted (green, blue, red and yellow lines) with

$f_{jk} = E_{Zj}(\varepsilon) + (-1)^{k+1}J(\varepsilon)$, where j denotes the qubit and k denotes the state of the other qubit. The data are fitted well using $J(\varepsilon) \propto e^{c_1\varepsilon}$, $E_{Z1}(\varepsilon) \propto e^{c_2\varepsilon}$, and $E_{Z2}(\varepsilon) \propto \varepsilon$. The fitted Zeeman energies of Q1 and Q2 are shown by the black lines. We observe that the Zeeman energy of Q1 has an exponential dependence towards the (0, 2) charge regime ($\varepsilon=0$). This observation can be explained by the electron delocalizing from D1 towards D2, which has a much higher Zeeman energy. **c,** Schematic of the colour-coded transitions that correspond to the resonances in **a** and **b**.



Extended Data Figure 9 | Decay of the decoupled CZ oscillations. The normalized spin-up probability of Q1 as a function of the total duration time 2τ of the two CZ gates in the DCZ sequence. The data are fitted using a sinusoid, $P_{|1\rangle} = 0.5\sin(2\pi J\tau) + 0.5$, with either a Gaussian (black line; $e^{-(2\tau/T_2)^2}$) or exponential (red line; $e^{-2\tau/T_2}$) decay. From these fits we find a decay time of $T_2 = 1.6 \mu\text{s}$. All error bars are 1σ from the mean.



Extended Data Figure 10 | Simulation of the Deutsch-Josza and Grover search algorithms using the DCZ gate. a, b, Two-spin probabilities as a function of the sequence time during the Deutsch-Josza algorithm (**a**) and the Grover search algorithm (**b**) for each function, using the decoupled

version of the two-qubit CZ gate (the DCZ gate). The solid lines show the outcome of the simulations, which include decoherence due to quasi-static charge noise and nuclear-spin noise. All error bars are 1σ from the mean.

Massive Dirac fermions in a ferromagnetic kagome metal

Linda Ye^{1*}, Mingu Kang^{1*}, Junwei Liu^{1†}, Felix von Cube^{2†}, Christina R. Wicker¹, Takehito Suzuki¹, Chris Jozwiak³, Aaron Bostwick³, Eli Rotenberg³, David C. Bell^{2,4}, Liang Fu¹, Riccardo Comin¹ & Joseph G. Checkelsky¹

The kagome lattice is a two-dimensional network of corner-sharing triangles¹ that is known to host exotic quantum magnetic states^{2–4}. Theoretical work has predicted that kagome lattices may also host Dirac electronic states⁵ that could lead to topological⁶ and Chern⁷ insulating phases, but these states have so far not been detected in experiments. Here we study the *d*-electron kagome metal Fe₃Sn₂, which is designed to support bulk massive Dirac fermions in the presence of ferromagnetic order. We observe a temperature-independent intrinsic anomalous Hall conductivity that persists above room temperature, which is suggestive of prominent Berry curvature from the time-reversal-symmetry-breaking electronic bands of the kagome plane. Using angle-resolved photoemission spectroscopy, we observe a pair of quasi-two-dimensional Dirac cones near the Fermi level with a mass gap of 30 millielectronvolts, which correspond to massive Dirac fermions that generate Berry-curvature-induced Hall conductivity. We show that this behaviour is a consequence of the underlying symmetry properties of the bilayer kagome lattice in the ferromagnetic state and the atomic spin–orbit coupling. This work provides evidence for a ferromagnetic kagome metal and an example of emergent topological electronic properties in a correlated electron system. Our results provide insight into the recent discoveries of exotic electronic behaviour in kagome-lattice antiferromagnets^{8–10} and may enable lattice-model realizations of fractional topological quantum states^{11,12}.

The kagome lattice (Fig. 1a) is a network with trihexagonal (3.6)² Archimedes tiling that has been studied extensively in the context of frustration-induced quantum-spin-liquid phases^{2–4}. In terms of electronic structure, simple tight-binding models on kagome lattices have long been known to yield unusual features, including dispersionless bands and Dirac points (Fig. 1b); the Dirac points appear in a manner similar to those in hexagonal graphene lattices¹³. Although such features have not previously been observed in experiments, theoretical interest has persisted and lead to several further predictions. Of particular interest are kagome networks in which time reversal symmetry is broken via ferromagnetism (Fig. 1c)^{5,7,11}, which has the effect of splitting the spin-degenerate Dirac bands (Fig. 1d). Further inclusion of spin–orbit coupling (Fig. 1e) yields various gapped phases (Fig. 1f) with integer⁷ or fractional^{11,12} topological invariants (Chern numbers). When the chemical potential is within the Dirac gap, the intrinsic anomalous Hall effect, which results from the integration of Berry curvature over the Brillouin zone, is quantized and in principle affords detection of the Chern number of the wavefunction of the system^{14,15}. More generally, with multiple bands and arbitrary positions of the chemical potential, such measurements can detect the Berry curvature that is concentrated by massive Dirac bands¹⁶.

Despite a broad theoretical understanding of electronic Berry-phase effects in ferromagnetic kagome models and extensive studies

of kagome insulators⁴, experimental realization of the former has been challenging, in part owing to the relative rarity of kagome materials. An approach to realizing metallic kagome networks in the hexagonal transition-metal stannides A_xSn_y (A = Mn, Fe or Co; *x*:*y* = 3:1, 3:2 or 1:1) has been reported¹⁷. As shown in Fig. 1g for A = Fe (studied here), starting from a single layer of a hexagonal close-packed structure of iron atoms, a kagome net emerges naturally by replacing a 2 × 2 sublattice (dashed cell) with tin atoms, resulting in an Fe₃Sn plane with an underlying iron kagome lattice.

Here we study the bilayer kagome compound Fe₃Sn₂ (space group *R*3*m*; hexagonal lattice constants *a* = 5.338 Å and *c* = 19.789 Å)—a structural variation of Fe_xSn_y that includes a stanene layer sandwiched between Fe₃Sn bilayers (Fig. 1h). In Fig. 1h we also show a corresponding transmission electron microscopy image of a (10 $\bar{1}$ 0) cross-section of a single crystal of Fe₃Sn₂, which reveals the Fe₃Sn and stanene layers. Previous studies¹⁷ have identified Fe₃Sn₂ as an unusual magnetic conductor with a high Curie temperature of *T*_C = 670 K. Although attention was originally focused on the zero-field spin structure¹⁸, recent studies have focused on the formation of skyrmion bubbles¹⁹ and a substantial anomalous Hall effect^{20,21} at finite field. The latter is particularly interesting in comparison with the structurally related antiferromagnets Mn₃Sn and Mn₃Ge, which were recently reported to have a large room-temperature anomalous Hall response^{8,9} and possible Weyl fermion states¹⁰.

Measurements of magnetization *M* as a function of magnetic induction *B* along the *c* axis (Fig. 2a) demonstrate that the system is a soft ferromagnet, with the saturation field and saturation magnetization *M*_s depending mildly on temperature *T*. The saturation magnetization *M*_s reaches approximately 1.9 μ_B per iron atom at low temperature *T* (where μ_B is the Bohr magneton; Fig. 2a, inset). The crystals exhibit high metallicity, with the residual resistivity ratio of ρ(300 K)/ρ(2 K) ≈ 25 (Extended Data Fig. 1) allowing characterization by electrical transport. The transverse resistivity in the kagome plane ρ_{yx}(*B*) (Fig. 2b) strongly reflects *M*(*B*)—a characteristic of the anomalous Hall effect^{20,21}. In ferromagnetic conductors it is conventional¹⁶ to express ρ_{yx}(*B*) in terms of contributions from the ordinary (Lorentz-force) Hall coefficient *R*₀ and the anomalous Hall coefficient *R*_s: ρ_{yx} = *R*₀*B* + *R*_s*M*; as shown in the inset of Fig. 2b, *R*₀ depends mildly on *T* (corresponding to 6 × 10²¹ electrons per cm³ at low *T*), whereas *R*_s is much larger but decreases with decreasing *T*.

To elucidate the role of the bilayer kagome lattice further, we examine the associated Hall conductivities in the kagome plane. The contributions to the total Hall conductivity σ_{xy} = σ_{xy}^N + σ_{xy}^A, where the superscript 'N' ('A') denotes the normal (anomalous) component, can be separated by using the field linearity of σ_{xy}^N in the low-Hall-angle limit (Methods). Although σ_{xy}^A is known to have contributions of both intrinsic (Berry curvature) and extrinsic (scattering) origin¹⁶, it has recently

¹Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ³Advanced Light Source, E. O. Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁴Center for Nanoscale Systems, Harvard University, Cambridge, Massachusetts 02138, USA. [†]Present addresses: Department of Physics, Hong Kong UST, Clear Water Bay, Hong Kong, China (J.L.); Hitachi High-Technologies Europe GmbH, Krefeld, Germany (F.v.C.).

*These authors contributed equally to this work.

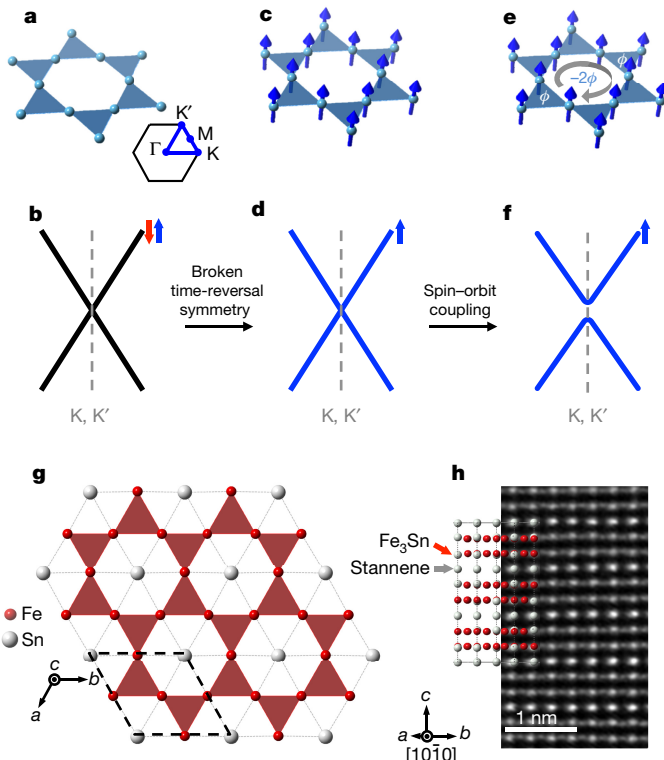


Figure 1 | The kagome structure and Fe_3Sn_2 . **a, b**, Structure of the kagome lattice (**a**) and the associated Dirac point in the nearest-neighbour tight-binding model (**b**), with the Brillouin zone shown in the inset. The band is degenerate, as denoted with red and blue spins. **c, d**, Ferromagnetic kagome lattice with broken time-reversal symmetry (moments in blue) (**c**) and the associated spin-polarized Dirac band with coupling between the magnetization and spin (**d**). **e, f**, Spin-orbit-coupled ferromagnetic kagome lattice with Berry phase ϕ accrued via hopping (**e**) and the corresponding gapped Dirac spectrum (**f**). **g**, The Fe_3Sn kagome plane in Fe_3Sn_2 , with the kagome network shown in red. **h**, Transmission electron microscopy cross-section of Fe_3Sn_2 and the corresponding Fe_3Sn and stanene layers viewed from the $[10\bar{1}0]$ direction.

been demonstrated that the insensitivity of the latter to thermal excitations allows the parameterization $\sigma_{xy}^A = f(\sigma_{xx,0})\sigma_{xx}^2 + \sigma_{xy}^{\text{int}}$, where $f(\sigma_{xx,0})$ is a function of the residual conductivity $\sigma_{xx,0}$, σ_{xx} is the conductivity and σ_{xy}^{int} is the intrinsic anomalous Hall conductivity²². Because σ_{xy}^{int} does not depend on the scattering rate, in a system with substantial Berry curvature, σ_{xy}^{int} is then the remnant σ_{xy}^A that is observed as $\sigma_{xx}^2 \rightarrow 0$ (Fig. 2c, top inset). Figure 2c demonstrates that $\sigma_{xy}^A(T)$ remains relatively unchanged from this remnant value at high temperature, until $T \approx 100$ K at which an upturn concomitant with increasing $\sigma_{xx}(T)$ is observed. This upturn is indicative of the onset of an extrinsic response, which is expected¹⁶ with the longer relaxation time in this range of σ_{xx} ; the subsequent σ_{xx}^2 scaling of the additional σ_{xy}^A (Fig. 2c, top inset) is also consistent with an extrinsic origin^{22,23}. The scattering-rate-independent value of σ_{xy}^A at high T persists, varying by about 10% down to $T = 2$ K ($158 \pm 16 \Omega^{-1} \text{cm}^{-1}$), which corresponds to approximately $0.27e^2/h$ per kagome bilayer, where e is the electronic charge and h is the Planck constant. We identify this contribution as σ_{xy}^{int} (Fig. 2c), with behaviour akin to that expected from a massive Dirac band²³.

These observations point to a substantial Berry-curvature contribution to the transport response in Fe_3Sn_2 in a geometry that samples the kagome planes (Fig. 2c, middle inset). We also measured the Hall response perpendicular to the kagome plane σ_{zx} (Fig. 2c, lower inset). This out-of-plane signal is much smaller (Fig. 2c), with the ratio $|\sigma_{zx}^A/\sigma_{xy}^A|$ being less than 10% at the highest T , indicating a large relative enhancement of the Berry curvature in the kagome plane.

To examine the origin of this Hall response further, we measured the electronic structure of Fe_3Sn_2 by using angle-resolved photoemission spectroscopy (ARPES). In Fig. 3a, b we show the experimental Fermi surface and energy-momentum dispersion, respectively, of the electronic bands along high-symmetry directions parallel to the kagome planes, measured at $T = 20$ K (see also Extended Data Fig. 2). A rich spectrum of electronic excitations with hexagonal symmetry is observed, consistent with the metallicity and crystallographic structure described above. More notably, linearly dispersing Dirac cones are observed at the corner points K and K' of the Brillouin zone. This spectrum, which is reminiscent of the electronic structure of graphene²⁴, is the long-sought realization of kagome-derived Dirac fermions⁵. These Dirac-like bands are shown in detail in the high-resolution energy-momentum section of the ARPES data across the K point in Fig. 3c (data are collected along the blue dashed line in Fig. 3a and then symmetrized in momentum about K); two Dirac cones, separated in energy but centred at K, are resolved. Hereafter, we focus on these bands and their role in generating Berry curvature.

The two-fold Dirac cones can also be identified in constant-energy contours (Fig. 3d). At the Fermi energy E_F (Fig. 3d, top layer), a pair of Dirac cones forms two electron pockets centred at K: a circular inner pocket and a trigonally warped outer pocket. Moving down from E_F each pocket shrinks, forming apparent Dirac points at binding energies of 70 meV (Fig. 3d, second layer) and 180 meV (Fig. 3d, bottom layer). At the midpoint energy (125 meV), the two Dirac cones cross and, within our experimental resolution, form a ring of Dirac points in the x - y momentum plane. The experimental electronic structure near the K point is therefore characterized by two energy-split ($\Delta E = 110$ meV) interpenetrating Dirac cones. This splitting is a natural consequence of the bilayer kagome structure, similarly to the case of multilayer graphene²⁴, whereas the exchange splitting due to magnetic order is expected²⁵ to be much larger (in excess of 2 eV). Photon-energy-dependent ARPES (Extended Data Fig. 4) reveals negligible variation in the Dirac bands as a function of out-of-plane momentum k_z , indicating quasi-two-dimensional (quasi-2D) bands confined to the iron kagome bilayer.

Having established the Dirac fermiology of Fe_3Sn_2 , we focus on the role of spin-orbit coupling and the possible mass acquisition of the Dirac bands. Inspection of the raw ARPES data reveals that the spectral intensity at the Dirac point is suppressed substantially (Extended Data Fig. 3d), which is more clearly visualized in the second derivative of the ARPES map (Fig. 3e). Analysis of the energy distribution curves displayed in Fig. 3f reveals a break between the upper and lower branches of the Dirac cone, which signals the opening of an energy gap Δ . A quantitative analysis performed by fitting the energy distribution curves with Gaussian peaks returns $\Delta = 30 \pm 5$ meV (Methods). This value is similar to that predicted previously for spin-orbit-coupled 3d transition metals in kagome lattices⁷, but smaller than that observed in magnetically doped topological insulators (about 50 meV)^{26,27} and in hydrogen-decorated graphene (at least 0.5 eV)²⁸.

The emergence of massive Dirac fermions in Fe_3Sn_2 can be understood as a combination of ferromagnetic splitting and spin-orbit coupling in the underlying kagome geometry. Motivated by the weak k_z dispersion observed in ARPES, we consider a stacked system of kagome layers. In Fig. 4a we show a perfect Fe_3Sn kagome layer and the corresponding Brillouin zone. The kagome layer has two-fold and three-fold rotational symmetries (C_{2x} and C_{3z} , respectively) that leave the K and K' points invariant and thus form point group D_3 . In the absence of spin-orbit coupling, the two-fold-degenerate crossing (Dirac) points at K and K' belong to a two-fold irreducible representation (E) and are therefore protected. As illustrated in Fig. 4b, a Dirac crossing can be observed at K in a tight-binding model H_K for nearest-neighbour hopping on the kagome sites:

$$H_K = \sum_{\langle ij \rangle} t c_i^\dagger c_j \quad (1)$$

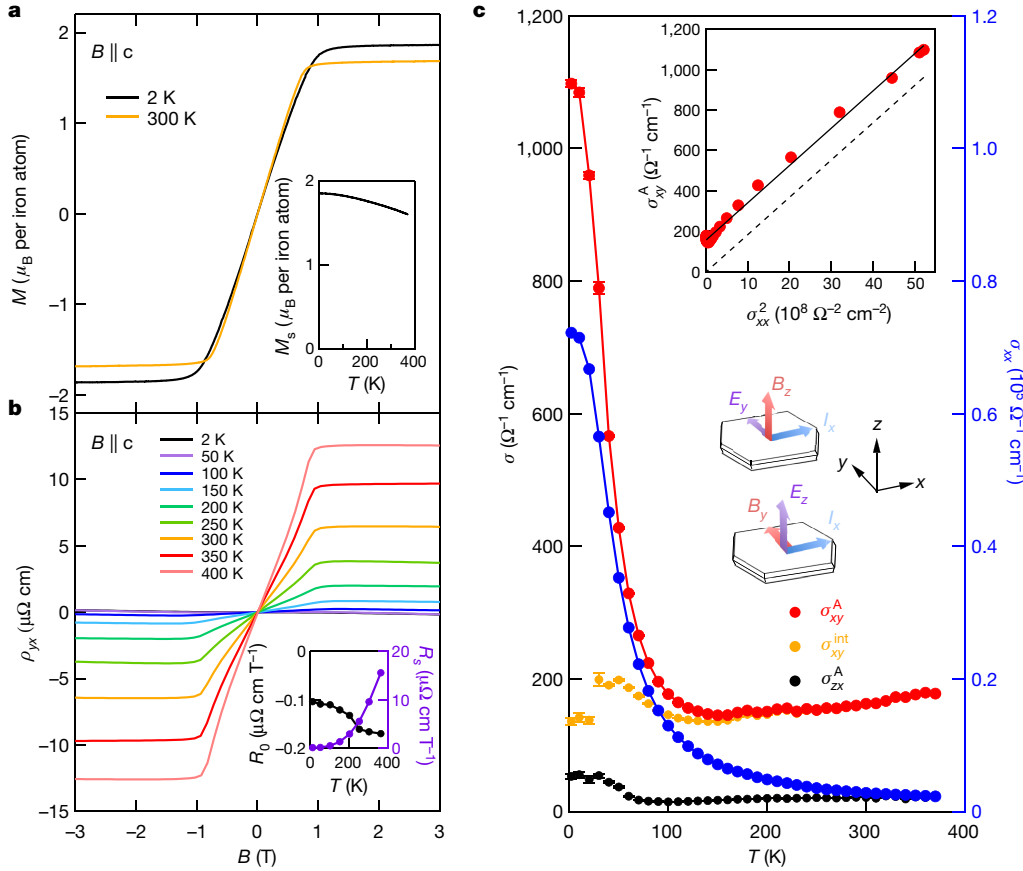


Figure 2 | Anomalous Hall response of Fe₃Sn₂. **a**, Magnetization M of Fe₃Sn₂ along the c axis as a function of magnetic induction B at room temperature ($T = 300$ K; orange) and low temperature ($T = 2$ K; black). The inset shows the saturation magnetization M_s (measured at 2 T) as a function of temperature T . **b**, Hall resistivity ρ_{yx} as a function of B . The inset shows the ordinary and anomalous Hall coefficients R_0 (black) and R_s (purple), respectively, as a function of T . **c**, Anomalous Hall conductivities

where $\langle ij \rangle$ indexes nearest-neighbour pairs, t is the hopping integral and c_j (c_j^\dagger) is the fermion annihilation (creation) operator, which is taken to be spin-polarized owing to exchange. The kagome bilayers in Fe₃Sn₂ (Fig. 4c) are tiled by triangles of two different bond lengths, 2.59 Å and 2.75 Å, as indicated by the red and blue shading. However, the combined unit of these kagome layers and the intervening stanene layer preserves the $\{C_{2v}, C_{3z}\}$ symmetry of the perfect kagome lattice and the Dirac points are thus protected by crystal symmetry in the absence of spin-orbit coupling. The additional layer degree of freedom further enriches the electronic structure. In particular, the ABA layer stacking of the structure in Fig. 4c gives rise to bonding-antibonding splitting²⁹, as seen in a simple tight-binding model with this additional hopping (Fig. 4d).

We next introduce Kane-Mele-type spin-orbit coupling H_{SOI} to the tight-binding model H_K , with

$$H_{\text{SOI}} = i \sum_{\langle ij \rangle} \lambda_{ij} (c_{i\uparrow}^\dagger c_{j\uparrow} - c_{i\downarrow}^\dagger c_{j\downarrow}) \quad (2)$$

where λ_{ij} represents the effect of spin-orbit coupling and \uparrow and \downarrow denote the spin quantum number³⁰. Writing $\lambda_{ij} = \lambda(\mathbf{E}_{ij} \times \mathbf{R}_{ij}) \cdot \mathbf{s}$, where λ is the spin-orbit coupling constant, \mathbf{E} is the electric field on the hopping path, \mathbf{R} is the hopping vector and \mathbf{s} represents the electron spin, for spin-polarized bands near K and K' with non-zero z polarization s_z , H_{SOI} effectively reduces to the Haldane term³¹. Accordingly, for the single-layer case (Fig. 4b), when E_F is positioned in the Dirac gap, the system enters a Chern insulating phase with quantized anomalous Hall conductance^{7,31}.

σ_{xy}^A (red, left axis) and σ_{zx}^A (black, left axis) in the x - y and z - x planes, respectively, along with the longitudinal conductivity σ_{xx} (blue, right axis) and estimated intrinsic Hall conductivity σ_{xy}^{int} (orange, left axis). The measurement configurations for σ_{xy} (top) and σ_{zx} (bottom) are shown in the lower inset; \mathbf{I} represents the charge current. The upper inset shows σ_{xy}^A plotted against σ_{xx}^2 ; the solid and dashed lines are the scaling curves (see text).

To connect with the Hall response, we construct a $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian near K and K' for the dual Dirac fermions and fit to the ARPES data (Fig. 4e, inset; Methods). We then calculate the contribution of the massive Dirac bands to the Hall response by integrating the Berry curvature of the filled states, which yields $\sigma_{xy}^{\text{calc}} = (0.31 \pm 0.05)e^2/h$ at E_F for a kagome bilayer (Fig. 4e), comparable to the observed value of $\sigma_{xy}^{\text{int}} = (0.27 \pm 0.03)e^2/h$ per bilayer. Remarkably, despite the simplicity of our model, the action of the quasi-2D massive Dirac fermions at K and K' largely accounts for the observed Hall response with the crystal viewed as a parallel network of bilayer kagome planes. However, there are limitations; for example, for a 2D model there is no contribution to the out-of-plane Hall response. We suggest instead that this out-of-plane response originates from the three-dimensional (3D) network of tin atoms and the associated k_z -dispersive bands near Γ (Methods). The relative smallness of σ_{zx}^A is then consistent with a minor contribution of tin-atom-derived bands to the overall Berry curvature. More generally, the model evidences the role of the concentration of Berry curvature in the quasi-2D massive Dirac bands, which have small E_F comparable to the spin-orbit coupling strength¹⁶. The robustness of the Hall response observed here is comparable to that of the Hall response that is driven by chiral antiferromagnetic order in Mn₃Sn (ref. 8) and Mn₃Ge (ref. 9); however, instead of originating from 3D Weyl nodes¹⁰, the Hall response observed here is driven by quasi-2D Dirac fermions in a ferromagnetic kagome network interleaved with stanene layers.

By combining electrical transport measurements, ARPES and theoretical analysis, this study provides a comprehensive proof of

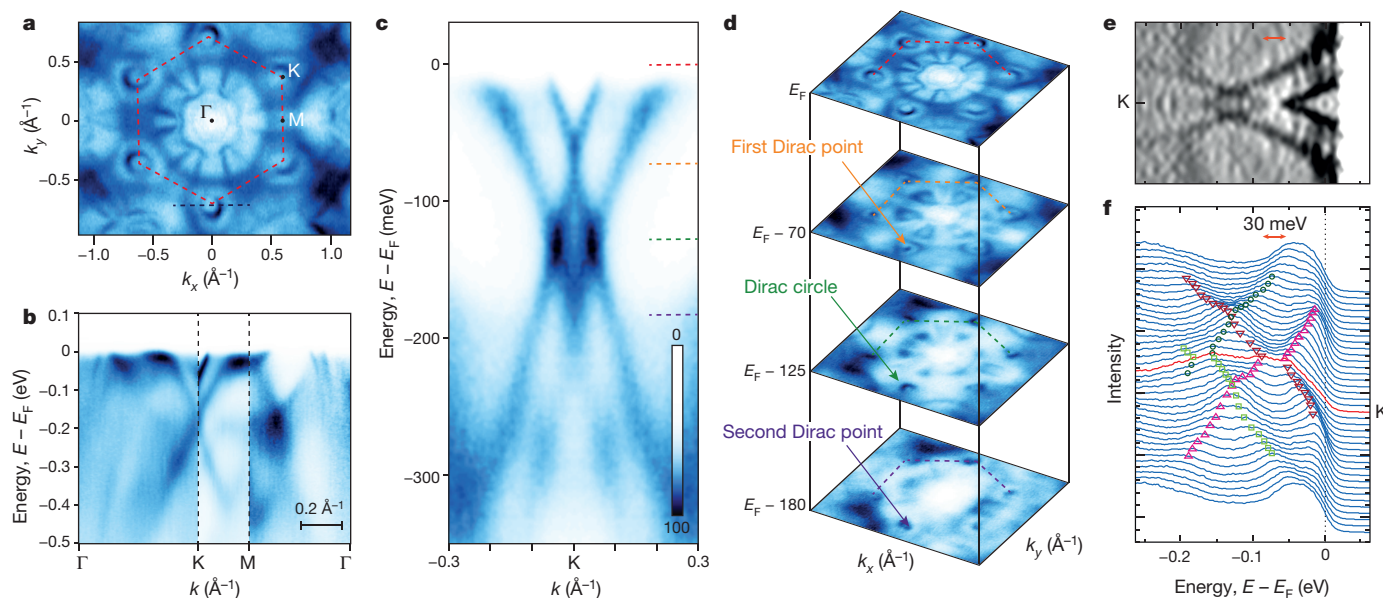


Figure 3 | Massive Dirac fermion at the zone corner of Fe_3Sn_2 .

a, Experimentally obtained Fermi surface of Fe_3Sn_2 . The hexagonal Brillouin zone (red dashed lines) and high-symmetry points (black dots) are marked. **b**, Experimentally determined band dispersion of Fe_3Sn_2 along the high-symmetry directions. **c**, High-resolution ARPES data, measured along the blue dashed line in **a** and then symmetrized with respect to K. The complete band dispersion is shown in Extended Data Fig. 3 (with a modified colour scale). **d**, Constant-energy maps at binding energies of 0 meV, 70 meV, 125 meV and 180 meV. Two electron pockets (top layer),

a first Dirac point (second layer), a Dirac circle (third layer) and a second Dirac point (bottom layer) can be clearly detected from the maps. The corresponding energies and Brillouin zone contours are marked with coloured dashed lines in **c** and **d**. **e**, **f**, The second-derivative plot (**e**) and the stack of energy distribution curves (**f**) across the Dirac points. Both panels share the momentum range and direction with **c**. The red double-headed arrow marks the discontinuity between the upper and lower branches of the Dirac cone. Coloured markers indicate the fitted band energies. All data were obtained using 92-eV photons.

principle for engineering band-structure singularities and topological phenomena in correlated systems. In particular, we realize lattice-driven^{6,7,30,31} topological 3d electronic bands, which we suggest exhibit the defining properties of a ferromagnetic kagome metal. Viewed in isolation, the bands near K can be considered to exhibit

a 2D ‘Chern gap’—a time-reversal-symmetry-broken topologically non-trivial phase that is intrinsic to stoichiometric materials and has a dominant contribution to the electrical response at temperatures of up to 300 K and above. To isolate these bands, as a step towards realizing high-temperature dissipationless modes¹⁵, we propose finding

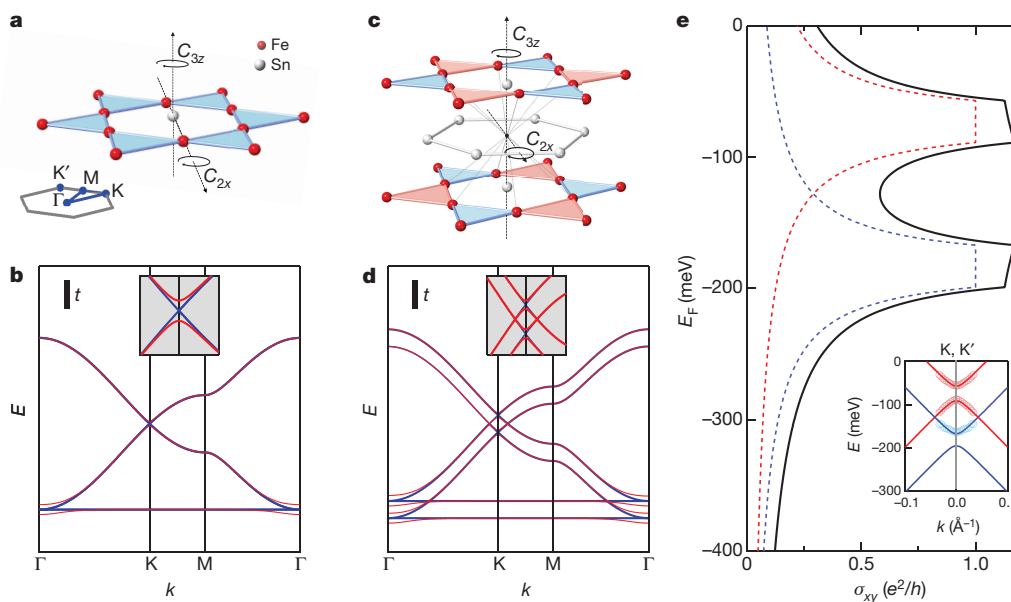


Figure 4 | Tight binding and hall conductivity of a kagome bilayer.

a, Two-fold (C_{2x}) and three-fold (C_{3z}) rotation-axis symmetry operations of a single Fe_3Sn kagome layer. **b**, Tight-binding band model of the single-layer kagome lattice with (red) and without (blue) spin-orbit coupling. The inset shows a magnified view of the (avoided) crossing near K. **c**, C_{2x} and C_{3z} symmetries of the ‘breathing’ kagome (illustrated by the red and blue inequivalent bonds) and stanene layers. **d**, Tight-binding band model of the double-layer kagome lattice with in-plane hopping t and inter-plane

hopping $0.3t$, with (red) and without (blue) spin-orbit coupling. The inset shows a magnified view of the double Dirac structure near K. The spin-orbit coupling strength $\lambda = 0.05t$ for **b** and **d**. **e**, Anomalous Hall conductivity σ_{xy} as a function of Fermi energy E_F from the $\mathbf{k} \cdot \mathbf{p}$ model (black solid curve; see text). The red and blue dashed curves represent the contributions from the upper and lower Dirac bands, respectively. The inset shows the fit (solid lines) of a massive Dirac dispersion to the ARPES results (circles) near K.

simplified structures with a single kagome layer per unit cell or pursuing 2D morphologies via thin films or exfoliation of related materials with van der Waals bonding. Furthermore, the frustrated hopping network of kagome lattices is predicted to support electronic bands (as in Fig. 4d) with non-zero Chern number and enhanced electronic correlation⁵. Searching for these bands by chemically doping Fe₃Sn₂, by using alternative structures and stoichiometries in A_xSn_y, or by using other spin–orbit-coupled 4d and 5d kagome metals are key to realizing new, magnetically driven fractionalized phases of matter^{11,12}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 September 2017; accepted 19 January 2018.

Published online 19 March 2018.

- O'Keeffe, M. & Hyde, B. G. *Crystal Structures. I. Patterns and Symmetry* Ch. 5 (Mineralogical Society of America, 1996).
- Sachdev, S. Kagome- and triangular-lattice Heisenberg antiferromagnets: ordering from quantum fluctuations and quantum-disordered ground states with unconfined bosonic spinons. *Phys. Rev. B* **45**, 12377–12396 (1992).
- Han, T.-H. *et al.* Fractionalized excitations in the spin-liquid state of a kagome-lattice antiferromagnet. *Nature* **492**, 406–410 (2012).
- Zhou, Y., Kanoda, K. & Ng, T.-K. Quantum spin liquid states. *Rev. Mod. Phys.* **89**, 025003 (2017).
- Mazin, I. I. *et al.* Theoretical prediction of a strongly correlated Dirac metal. *Nat. Commun.* **5**, 4261 (2014).
- Guo, H.-M. & Franz, M. Topological insulator on the kagome lattice. *Phys. Rev. B* **80**, 113102 (2009).
- Xu, G., Lian, B. & Zhang, S.-C. Intrinsic quantum anomalous Hall effect in the kagome lattice Cs₂LiMn₃F₁₂. *Phys. Rev. Lett.* **115**, 186802 (2015).
- Nakatsuji, S., Kiyohara, N. & Higo, T. Large anomalous Hall effect in a non-collinear antiferromagnet at room temperature. *Nature* **527**, 212–215 (2015).
- Nayak, A. K. *et al.* Large anomalous Hall effect driven by a nonvanishing Berry curvature in the noncollinear antiferromagnet Mn₃Ge. *Sci. Adv.* **2**, e1501870 (2016).
- Kuroda, K. *et al.* Evidence for magnetic Weyl fermions in a correlated metal. *Nat. Mater.* **16**, 1090–1095 (2017).
- Tang, E. & Wen, X.-G. High-temperature fractional quantum Hall states. *Phys. Rev. Lett.* **106**, 236802 (2011).
- Bergholtz, E. J., Liu, Z., Trescher, M., Moessner, R. & Udagawa, M. Topology and interactions in a frustrated slab: tuning from Weyl semimetals to C > 1 fractional Chern insulators. *Phys. Rev. Lett.* **114**, 016806 (2015).
- Wallace, P. R. The band theory of graphite. *Phys. Rev.* **71**, 622–634 (1947).
- Thouless, D. J., Kohmoto, M., Nightingale, M. P. & den Nijs, M. Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
- Chang, C.-Z. *et al.* Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167–170 (2013).
- Nagaosa, N., Sinova, J., Onoda, S., MacDonald, A. H. & Ong, N. P. Anomalous Hall effect. *Rev. Mod. Phys.* **82**, 1539–1592 (2010).
- Giefers, H. & Nicol, M. High pressure X-ray diffraction study of all FeSn intermetallic compounds and one FeSn solid solution. *J. Alloys Compd.* **422**, 132–144 (2006).
- Le Caër, G., Malaman, B. & Roques, B. Mössbauer effect study of Fe₃Sn₂. *J. Phys. F* **8**, 323–336 (1978).
- Hou, Z. *et al.* Observation of various and spontaneous magnetic Skyrmionic bubbles at room temperature in a frustrated kagome magnet with uniaxial magnetic anisotropy. *Adv. Mater.* **29**, 1701144 (2017).
- Kida, T. *et al.* The giant anomalous Hall effect in the ferromagnet Fe₃Sn₂—a frustrated kagome metal. *J. Phys. Condens. Matter* **23**, 112205 (2011).
- Wang, Q., Sun, S., Zhang, X., Pang, F. & Lei, H. Anomalous Hall effect in a ferromagnetic Fe₃Sn₂ single crystal with a geometrically frustrated Fe bilayer kagome lattice. *Phys. Rev. B* **94**, 075135 (2016).
- Tian, Y., Ye, L. & Jin, X. Proper scaling of the anomalous Hall effect. *Phys. Rev. Lett.* **103**, 087206 (2009).
- Shitade, A. & Nagaosa, N. Anomalous Hall effect in ferromagnetic metals: role of phonons at finite temperature. *J. Phys. Soc. Jpn* **81**, 083704 (2012).
- Kim, K. S. *et al.* Coexisting massive and massless Dirac fermions in symmetry-broken bilayer graphene. *Nat. Mater.* **12**, 887–892 (2013).
- Sales, B. C., Saporov, B., McGuire, M. A., Singh, D. J. & Parker, D. S. Ferromagnetism of Fe₃Sn and alloys. *Sci. Rep.* **4**, 7024 (2014).
- Chen, Y. L. *et al.* Massive Dirac fermion on the surface of a magnetically doped topological insulator. *Science* **329**, 659–662 (2010).
- Xu, S. Y. *et al.* Hedgehog spin texture and Berry's phase tuning in a magnetic topological insulator. *Nat. Phys.* **8**, 616–622 (2012).
- Balog, R. *et al.* Bandgap opening in graphene induced by patterned hydrogen adsorption. *Nat. Mater.* **9**, 315–319 (2010).
- Ishii, Y., Harima, H., Okamoto, Y., Yamaura, J. & Hiroi, Z. YCr₆Ge₆ as a candidate compound for a kagome metal. *J. Phys. Soc. Jpn* **82**, 023705 (2013).
- Kane, C. L. & Mele, E. J. Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95**, 226801 (2005).
- Haldane, F. D. M. Model for a quantum Hall effect without Landau levels: condensed-matter realization of the “parity anomaly”. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).

Acknowledgements We are grateful to X.-G. Wen and E. Tang for discussions. This research was funded in part by the Gordon and Betty Moore Foundation EPIQS Initiative, grant GBMF3848 to J.G.C. and NSF grant DMR-1554891. L.Y., J.L. and F.v.C. acknowledge support by the STC Center for Integrated Quantum Materials, NSF grant number DMR-1231319. L.Y. acknowledges support by the Tsinghua Education Foundation. M.K. acknowledges a Samsung Scholarship from the Samsung Foundation of Culture. This research used resources of the Advanced Light Source, which is a DOE Office of Science User Facility under contract number DE-AC02-05CH11231. A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by NSF cooperative agreement number DMR-1157490, the State of Florida and the US Department of Energy.

Author Contributions L.Y., T.S. and C.R.W. grew the single crystals. L.Y. characterized the materials, performed the transport and magnetic measurements and analysed the resultant data. M.K., C.J., A.B. and E.R. performed the ARPES experiment and analysed the resultant data. J.L. and L.Y. performed the theoretical calculations. F.v.C. and D.C.B. performed the electron microscopy study. All authors contributed to writing the manuscript. L.F., R.C. and J.G.C. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.C. (rcomin@mit.edu) or J.G.C. (checkelsky@mit.edu).

Reviewer Information Nature thanks E. Bergholtz, B. Lake and O. Rader for their contribution to the peer review of this work.

METHODS

Single-crystal growth. Single crystals of Fe_3Sn_2 were grown using an I_2 -catalysed reaction. A stoichiometric ratio of Fe and Sn powders was sealed in a quartz tube with approximately 1% I_2 by mass and kept in a horizontal three-zone furnace with a temperature gradient from 750 °C to 650 °C for five weeks followed by water quenching to stabilize the Fe_3Sn_2 phase. Hexagonal, plate-like crystals of sub-millimetre size (Extended Data Fig. 1a, inset) formed near the high-temperature region as has been reported previously³² for Fe_3Ge . The hexagonal surfaces were confirmed as (0001) kagome planes by using single-crystal X-ray diffraction.

Magnetization measurements. Magnetization measurements were performed using a commercial superconducting quantum interference device (SQUID) magnetometer with the field oriented along the c axis and in the a - b plane. Demagnetization corrections were performed for all measurements. The measured saturation moment ($1.9\mu_B$ per Fe atom) is consistent with previous reports^{18,19,33}.

Transport measurements. Four probe transport measurements were performed for longitudinal and Hall resistivity in a commercial cryostat with a superconducting magnet. High-field transport measurements in fields of up to 31 T were performed in a He-3 cryostat at Cell-9 of the National High Magnetic Field Laboratory. For measurements in the kagome plane, the field was applied along the [0001] direction with current and voltages in the kagome plane. For Hall measurements perpendicular to the kagome plane, the magnetic field and current were applied orthogonally in the kagome plane and the out-of-plane voltage is measured. The choice of coordination for in-plane and out-of-plane Hall measurements is such that, for the ordinary Hall effect of holes, $\sigma_{xy} > 0$ when the field is applied along the $+z$ direction and $\sigma_{xx} > 0$ ($\sigma_{xx} < 0$) when the field is applied along the $+y$ direction. Electrical signals were detected using standard AC lock-in techniques with a typical current density of 10 A cm^{-2} . To correct for contact misalignment, the measured longitudinal and transverse voltages were field-symmetrized and field-antisymmetrized, respectively. Demagnetization corrections were performed for all measurements.

Scanning transmission electron microscopy (STEM). STEM experiments were conducted at a probe-corrected STEM (JEOL ARM) operated at an acceleration voltage of 200 kV. Fe_3Sn_2 samples were prepared by a Helios focused-ion beam (FEI), operated at an acceleration voltage of 30 kV for the gallium beam during lift-out and of 2 kV during polishing. Additional polishing was performed at 1 kV and 0.5 kV with a NanoMill (Fischione). At both acceleration voltages, samples were polished for 20 min on each side.

Angle-resolved photoemission spectroscopy (ARPES). ARPES experiments were conducted at the Microscopic and Electronic Structure Observatory (MAESTRO) at beamline 7 (main data) and at the meV Resolution Soft X-ray Inelastic Scattering Beamline (MERLIN) at beamline 4 (preliminary measurement) of the Advanced Light Source. The two ARPES endstations are equipped with R4000 hemispherical electron analysers (VG scienta, Sweden). Data in Fig. 3 and Extended Data Fig. 3 were collected at 20 K with a photon energy of 92 eV, which maximizes the ARPES spectral weight of the Dirac bands. The photon-energy-dependent measurement was conducted from 45 eV to 120 eV (Extended Data Fig. 4). Energy and momentum resolutions were better than 15 meV and 0.01 \AA^{-1} , respectively. Fe_3Sn_2 samples were cleaved in the ultrahigh-vacuum chamber with a base pressure of better than 4×10^{-11} torr. All of the data were collected within 8 h after cleaving to minimize the effect of sample degradation. Six different samples from various growth batches were analysed to confirm the consistency of results.

Longitudinal electrical transport. The resistivity in the kagome plane of sample C1 as a function of temperature, $\rho(T)$, is shown in Extended Data Fig. 1a. A metallic response is seen at all T , with a residual resistivity ratio of $\rho(300\text{ K})/\rho(2\text{ K}) = 25$. The magnetoresistance for magnetic induction B normal to and within the kagome plane is shown in Extended Data Fig. 1b, c, respectively. In both figures, the electrical current I is perpendicular to the applied field. For $B \parallel c$ (Extended Data Fig. 1b), we observe a non-monotonic response below the saturation field ($B \leq 1.2\text{ T}$), which may reflect a transition through a skyrmion bubble phase, as was reported recently¹⁹. This structure is absent for $B \perp c$ (Extended Data Fig. 1c). In addition, for both $B \parallel c$ and $B \perp c$, a negative linear magnetoresistance characteristic of magnon suppression³⁴ is observed at high T .

Analysis of the Hall effect. In a ferromagnetic metal, the total Hall conductivity is composed of contributions from the normal Hall effect ('N') induced by Lorentz force and the anomalous Hall effect ('A'): $\sigma_{xy} = \sigma_{xy}^N + \sigma_{xy}^A$. In the limit of small Hall angle ($\Theta_H \equiv \rho_{yx}/\rho_{xx} \ll 1$), we have $\sigma_{xy} \approx B/(ne\rho_{xx}^2)$, where n is the carrier density and e is the electronic charge. Given the relatively small magnetoresistance, we extract σ_{xy}^N from a linear fit of $\sigma_{xy}(B)$ for fields above saturation (1.7 T to 5 T), with the intercept returning σ_{xy}^A (Extended Data Fig. 2a). That the condition $\Theta_H \ll 1$ is satisfied is confirmed to high field, with $\Theta_H \ll 0.04$ for fields up to 30 T (Extended Data Fig. 2b).

We measured the anomalous Hall effect at temperature $T = 300\text{ K}$ in multiple samples. We find a consistently enhanced anomalous Hall conductivity in the kagome plane σ_{xy}^A relative to out of the kagome plane σ_{xy}^A . For samples C1, C4 and C5, the observed value of σ_{xy}^A is $163.6\text{ }\Omega^{-1}\text{ cm}^{-1}$, $179.1\text{ }\Omega^{-1}\text{ cm}^{-1}$ and $138.8\text{ }\Omega^{-1}\text{ cm}^{-1}$, respectively. For samples C2, C4, C5 and C6, the observed value of σ_{xy}^A is $20.5\text{ }\Omega^{-1}\text{ cm}^{-1}$, $22.0\text{ }\Omega^{-1}\text{ cm}^{-1}$, $55.6\text{ }\Omega^{-1}\text{ cm}^{-1}$ and $53.7\text{ }\Omega^{-1}\text{ cm}^{-1}$, respectively. The T dependence for C1 and C4 is shown in Fig. 2.

Energy-momentum dispersion along high-symmetry directions. Extended Data Fig. 3a, e shows the experimental Fermi surface of Fe_3Sn_2 obtained from different orientations of the crystal axis with respect to the photoelectron emission plane. Extended Data Fig. 3b–d, f, g shows the experimental band dispersion of Fe_3Sn_2 along high-symmetry directions. Despite the occurrence of a complex-matrix-element effect near the K point, the structure of the two interpenetrating Dirac cones is clearly visible in all momentum directions probed in this experiment. In Extended Data Fig. 3h, energy distribution curves at different Brillouin-zone corner points are shown. All energy distribution curves show a consistent two-peak structure near the first Dirac point, signalling a gap opening. The velocity of the Dirac fermions v_D is found to be isotropic in the kagome plane, with magnitude $v_D = (1.76 \pm 0.11) \times 10^5\text{ m s}^{-1}$, comparable to that observed recently in iron pnictide³⁵ and selenide³⁶ superconductors, but lower than in graphene^{37,38}, possibly reflecting the correlated character of the Fe-3d states.

Photon-energy-dependent ARPES. Photon-energy-dependent ARPES probes the variation of the band structure along the momentum direction perpendicular to the sample surface (the k_z direction in our geometry)³⁹. The experiment here was conducted by varying the photon energy from 55 eV to 120 eV. The ARPES spectral weight of the localized d electrons is suppressed at low photon energy, so we report only the results from 90 eV to 120 eV. Considering the large c -axis lattice constant ($c = 19.8\text{ \AA}$), the photon energy range presented here spans more than three complete Brillouin zones. The dependence on photon energy varies substantially between different bands. For example, the dispersion of the innermost electron pocket at Γ (Extended Data Fig. 4b) depends critically on the photon energy (k_z), whereas the dispersion of Dirac bands shows negligible dependence on k_z . The latter finding demonstrates the pronounced 2D nature of the Dirac fermions that we observed. We ascribe the highly dispersive bands centred at Γ to the 3D network of Sn atoms within the crystal structure.

Tight-binding models. We use a simple 2D tight-binding model to capture the symmetry-protected Dirac nodes at K and K' in the single-layer (Fig. 4a) and bilayer (Fig. 4c) kagome lattice. Starting with the former, in the absence of spin-orbit coupling, the nearest-neighbour tight-binding model H_K (equation (1)) yields Dirac bands and a dispersionless band (Fig. 4b). For the bilayer kagome lattice, we include additional hopping $t_0 = 0.3t$ between the vertically displaced sites on each kagome layer, which introduces a layer splitting of the Dirac states in energy (Fig. 4d)²⁹. We examine the effect of the spin-orbit interaction by adding a Kane-Mele-type term³⁰ H_{SOI} (equation (2)) in the leading-order nearest-neighbour hopping to the tight-binding model H_K with strength given by $\lambda(E_{ij} \times \mathbf{R}_{ij}) \cdot \mathbf{s}$. The magnetic moments in Fe_3Sn_2 have been proposed to be subjected to spin reorientation and microscopic domain formation^{19,33}; contributions to λ_{ij} arise from orthogonality of the hopping path and local electric-field and spin directions in each domain. In the simplified hopping model, these contributions are represented by an in-plane electric field and s_z , with $\lambda_{ij} = 0.05t$ in Fig. 4b, d. Similar Hamiltonians^{40,41} have been used in the spin sector for insulating materials with kagome structures^{40–43}.

Model calculation of Hall conductivity. To connect with the Hall response, we first construct a $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian near K and K' for the dual Dirac fermions observed in the ARPES spectra:

$$H_D = [\hbar v_F(k_x\sigma_y - k_y\sigma_x)] \otimes I + E_0\tau_x + m\sigma_z \quad (3)$$

where σ_i ($i = x, y, z$) are the Pauli matrices of pseudospin for each Dirac band, E_0 is the energy splitting of the Dirac bands described by the Pauli matrix τ_x , and $m = \Delta/2$ is the Dirac mass. To obtain the band parameters, we fit the observed dispersion $E(k)$ to the massive Dirac model $E_{\pm}^i(k) = \pm \sqrt{(\hbar v_D k)^2 + (\Delta/2)^2} + E_0^i$, where E_0^i is the energy offset of the upper ($i = 1$) and lower ($i = 2$) Dirac bands from E_F . As shown in the inset of Fig. 4e, a satisfactory fit is found with $v_D = (1.85 \pm 0.15) \times 10^5\text{ m s}^{-1}$, $\Delta = 32 \pm 3\text{ meV}$, $E_0^1 = -73 \pm 5\text{ meV}$ and $E_0^2 = -182 \pm 5\text{ meV}$. The Dirac band centred at 73 meV can be analysed reliably, whereas matrix-element effects and the proximity of neighbouring bands interfere with the intensity distribution of the lower Dirac point. Applying the same set of Fermi velocity and mass for the upper and lower Dirac bands, we then calculate the contribution of the massive Dirac bands to the Hall response by integrating the Berry curvature over the filled states described by equation (3) as follows.

The Hall conductivity σ_{xy} can be considered a geometric quantity that characterizes the mapping from the k_x - k_y momentum plane to the two-component

Dirac band structure⁴⁴. For a general two-level Hamiltonian $\mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma}$, (where $\boldsymbol{\sigma} = \{\sigma_x, \sigma_y, \sigma_z\}$ and $\mathbf{d}(\mathbf{k})$ is a generic vector), the wave functions may be represented as points on the unit Bloch sphere by $\pm \hat{\mathbf{d}}(\mathbf{k}) = \mathbf{d}(\mathbf{k})/|\mathbf{d}(\mathbf{k})|$, where the \pm denotes the two eigenstates of the Hamiltonian at a given \mathbf{k} . σ_{xy} then takes the form⁴⁴

$$\sigma_{xy} = \frac{e^2}{h} \iint_{\text{filled states}} \hat{\mathbf{d}} \cdot (\partial_x \hat{\mathbf{d}} \times \partial_y \hat{\mathbf{d}}) dk_x dk_y \quad (4)$$

where the integrand (the Berry curvature) can be seen as the Jacobian associated with mapping the k_x - k_y plane to the unit sphere. Therefore, σ_{xy} is proportional to the total area covered on the unit sphere by the filled states. Also, because the integrand in equation (4) is a pseudoscalar, the states represented by $\pm \hat{\mathbf{d}}$ have opposite contributions to σ_{xy} .

The above formulation for σ_{xy} of a single Dirac fermion is illustrated in Extended Data Fig. 5 for the gapless and gapped cases. For the former (Extended Data Fig. 5a), the wavefunction of the Dirac fermions is confined to the equator of the Bloch sphere and therefore $\sigma_{xy} = 0$. Extended Data Fig. 5b shows the case in which the lower branch of the massive Dirac cone is filled. In this case, the occupied states span the lower hemisphere and yield $\sigma_{xy} = 0.5e^2/h$. When the upper branch of the gapped Dirac fermion is partially filled (Extended Data Fig. 5c), a contribution of opposite sign appears. The resulting σ_{xy} for a single Dirac fermion is therefore

$$\sigma_{xy} = \begin{cases} \frac{e^2}{h} \frac{S}{4\pi} & \text{for } S \leq 2\pi \\ \frac{e^2}{h} \frac{4\pi - S}{4\pi} & \text{for } S \geq 2\pi \end{cases}$$

where S is the total area of the filled states on the Bloch sphere; this is shown in Extended Data Fig. 5d as a function of energy E normalized by $\hbar v_F$ (where v_F is the Fermi velocity).

The above formulation is consistent with that expected from the semi-classical Boltzmann equation and Kubo formalism for a single massive Dirac fermion⁴⁵:

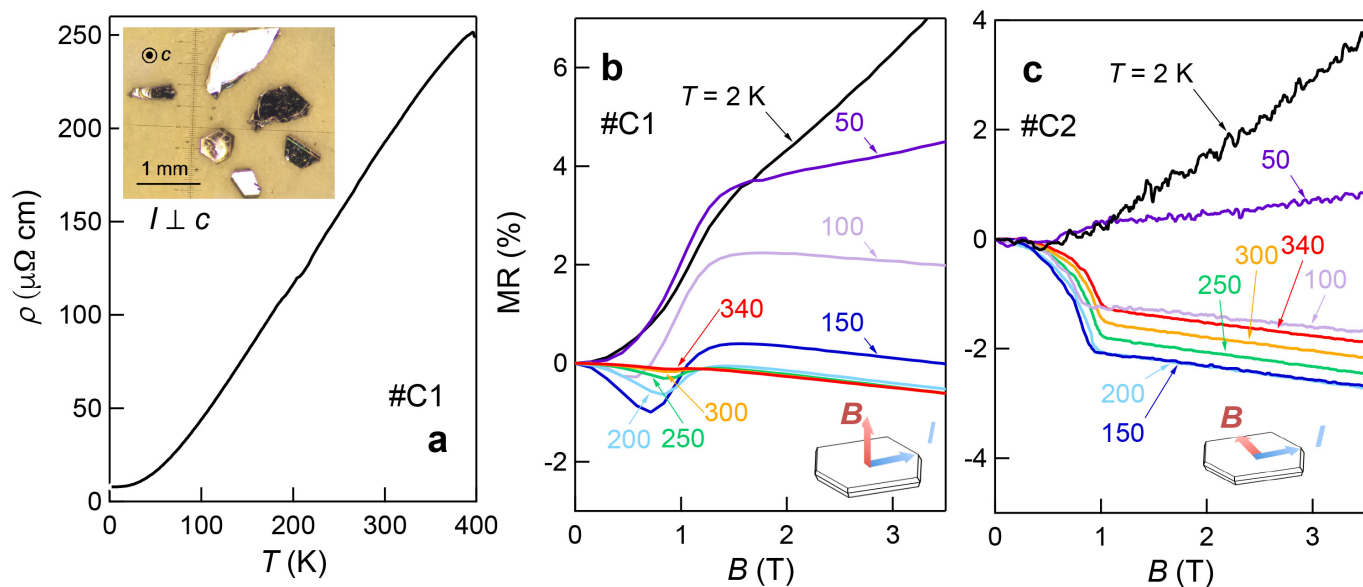
$$\sigma_{xy} = \frac{e^2}{h} \frac{\Delta/2}{\sqrt{(\Delta/2)^2 + (\hbar v_F k_F)^2}} \quad (5)$$

where Δ is the size of the Dirac gap and k_F is the Fermi wavevector. From equation (5) and Extended Data Fig. 5c, we see that σ_{xy} is maximized to $e^2/(2h)$ when E_F is within the gap and drops outside the gap with a long tail^{46,47}. Importantly, the massive Dirac fermions at the K and K' valleys are related by inversion symmetry and therefore contribute similarly to the Berry curvature; the contribution from equation (5) is therefore doubled. With the experimental fits to

equation (3) and adding up the contributions from the upper and lower Dirac points to the anomalous Hall conductivity described by equations (4) and (5), we obtain the energy-dependent 2D Hall conductance σ_{xy}^A (Fig. 4e). The Hall conductivity at the Fermi level evaluates to $\sigma_{xy}(E_F) = (0.31 \pm 0.05)e^2/h$. The uncertainty within this model arises from fitting the experimental band parameters near K; developing models that use the complete electronic structure and Berry curvature to compare to the experimental results is an important future direction.

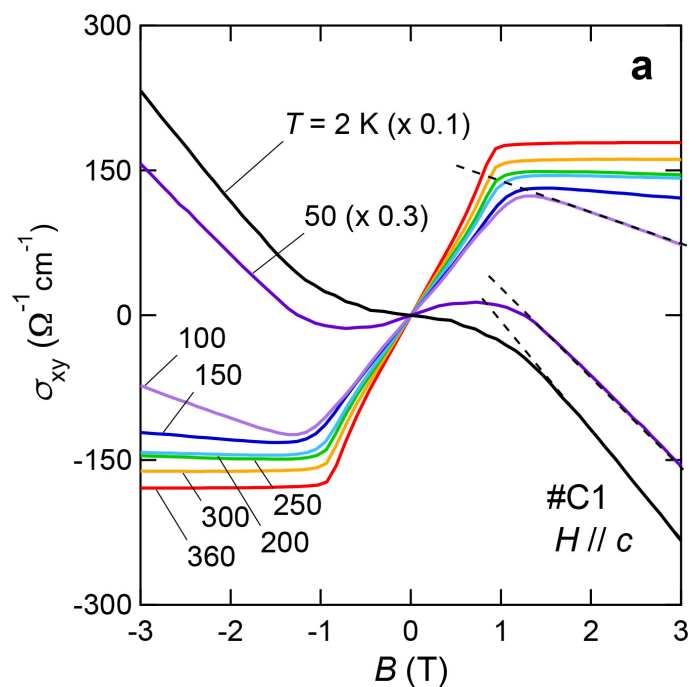
Data availability. The data that support the findings of this study are available from the corresponding authors on reasonable request.

32. Drijver, J. W., Sinnema, S. G. & van der Woude, F. Magnetic properties of hexagonal and cubic Fe₃Ge. *J. Phys. F* **6**, 2165–2177 (1976).
33. Fenner, L. A., Dee, A. A. & Wills, A. S. Non-collinearity and spin frustration in the itinerant kagome ferromagnet Fe₃Sn₂. *J. Phys. Condens. Matter* **21**, 452202 (2009).
34. Raquet, B., Viret, M., Sondergard, E., Cespedes, O. & Mamy, R. Electron-magnon scattering and magnetic resistivity in 3d ferromagnets. *Phys. Rev. B* **66**, 024433 (2002).
35. Richard, P. *et al.* Observation of Dirac cone electronic dispersion in BaFe₂As₂. *Phys. Rev. Lett.* **104**, 137001 (2010).
36. Tan, S. Y. *et al.* Observation of Dirac cone band dispersions in FeSe thin films by photoemission spectroscopy. *Phys. Rev. B* **93**, 104513 (2016).
37. Bostwick, A., Ohta, T., Seyller, T., Horn, K. & Rotenberg, E. Quasiparticle dynamics in graphene. *Nat. Phys.* **3**, 36–40 (2007).
38. Nevius, M. S. *et al.* Semiconducting graphene from highly ordered substrate interactions. *Phys. Rev. Lett.* **115**, 136802 (2015).
39. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
40. Onose, Y. *et al.* Observation of the magnon Hall effect. *Science* **329**, 297–299 (2010).
41. Chisnell, R. *et al.* Topological magnon bands in a kagome lattice ferromagnet. *Phys. Rev. Lett.* **115**, 147201 (2015).
42. Inami, T., Nishiyama, M., Maegawa, S. & Oka, Y. Magnetic structure of the kagome lattice antiferromagnet potassium jarosite KFe₃(OH)₆(SO₄)₂. *Phys. Rev. B* **61**, 12181–12186 (2000).
43. Hiroi, Z. *et al.* Spin-1/2 kagome-like lattice in volborthite Cu₃V₂O₇(OH)₂ · 2H₂O. *J. Phys. Soc. Jpn* **70**, 3377–3384 (2001).
44. Qi, X. L., Wu, Y.-S. & Zhang, S.-C. Topological quantization of the spin Hall effect in two-dimensional paramagnetic semiconductors. *Phys. Rev. B* **74**, 085308 (2006).
45. Sinitsyn, N. A., MacDonald, A. H., Jungwirth, T., Dugaev, V. K. & Sinova, J. Anomalous Hall effect in a two-dimensional Dirac band: the link between the Kubo-Streda formula and the semiclassical Boltzmann equation approach. *Phys. Rev. B* **75**, 045315 (2007).
46. Haldane, F. D. M. Berry curvature on the Fermi surface: anomalous Hall effect as a topological Fermi-liquid property. *Phys. Rev. Lett.* **93**, 206602 (2004).
47. Fang, Z. *et al.* The anomalous Hall effect and magnetic monopoles in momentum space. *Science* **302**, 92–95 (2003).

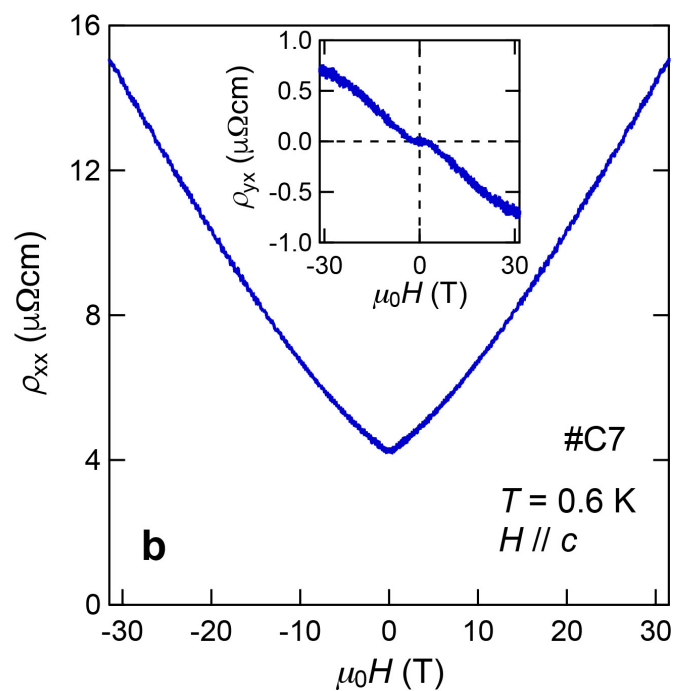


Extended Data Figure 1 | Metallic transport in Fe_3Sn_2 . **a**, Resistivity ρ as a function of temperature T in the kagome plane for Fe_3Sn_2 sample C1. The inset shows a photograph of Fe_3Sn_2 single crystals.

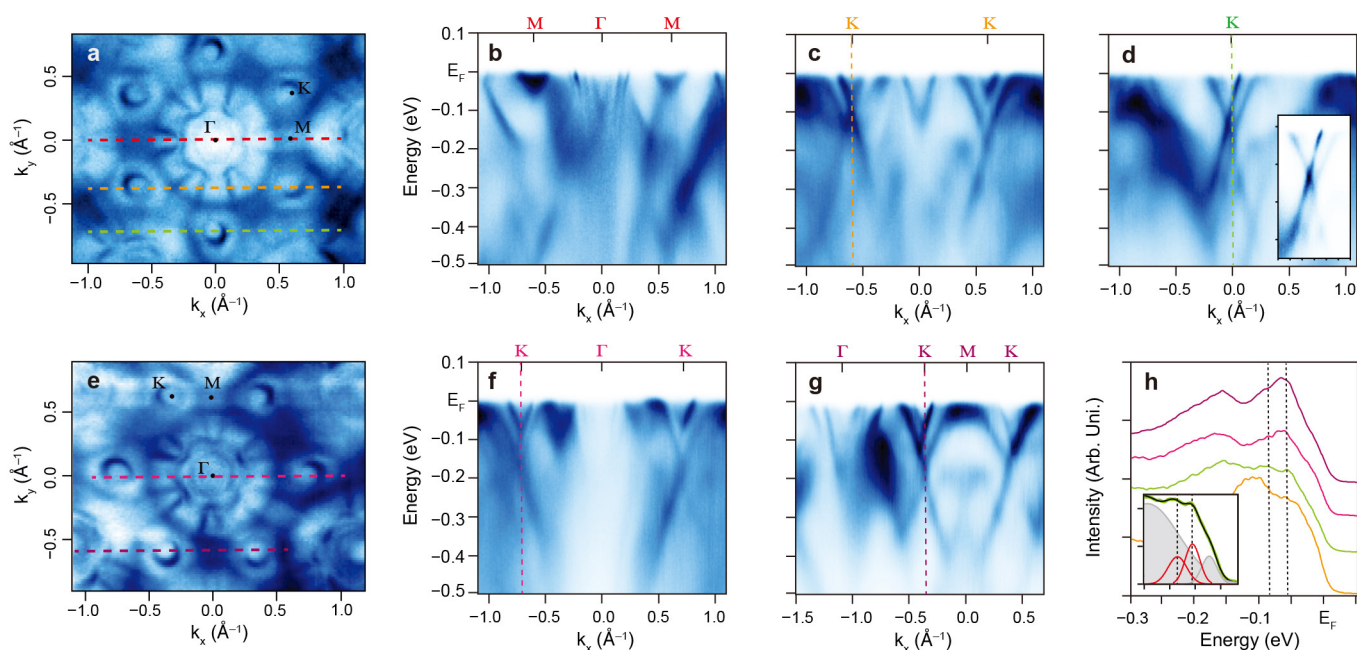
b, c, Magnetoresistance (defined as $\text{MR} = [\rho_{xx}(B) - \rho_{xx}(0)]/\rho_{xx}(0)$) at selected T with B applied perpendicular (**b**) or parallel (**c**) to the kagome plane and $B \perp I$ (schematics of the configurations are shown as insets).



Extended Data Figure 2 | Extracting anomalous Hall conductivity and high-field transport. a, In-plane Hall conductivity σ_{xy} as a function of magnetic induction B at selected temperatures. Dashed lines represent the

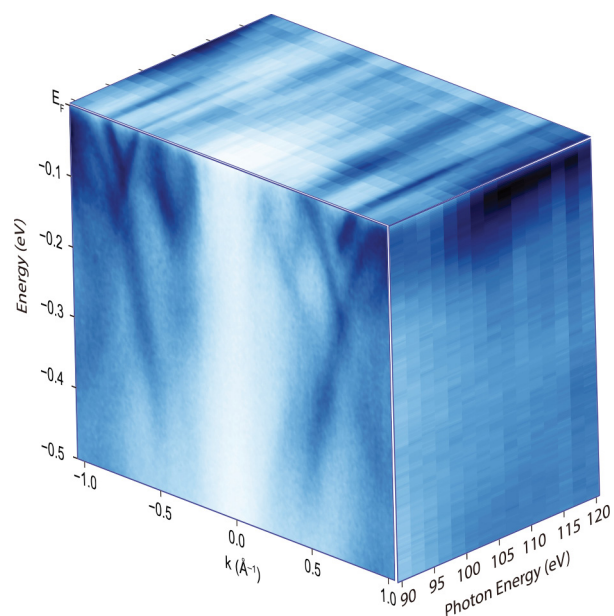


linear fit to σ_{xy}^N . The data at 2 K and 50 K have been scaled by the factors shown for clarity. **b,** Magnetoresistance (main panel) and Hall effect (inset) of Fe_3Sn_2 with applied magnetic field $\mu_0 H \parallel c$ up to 31 T.



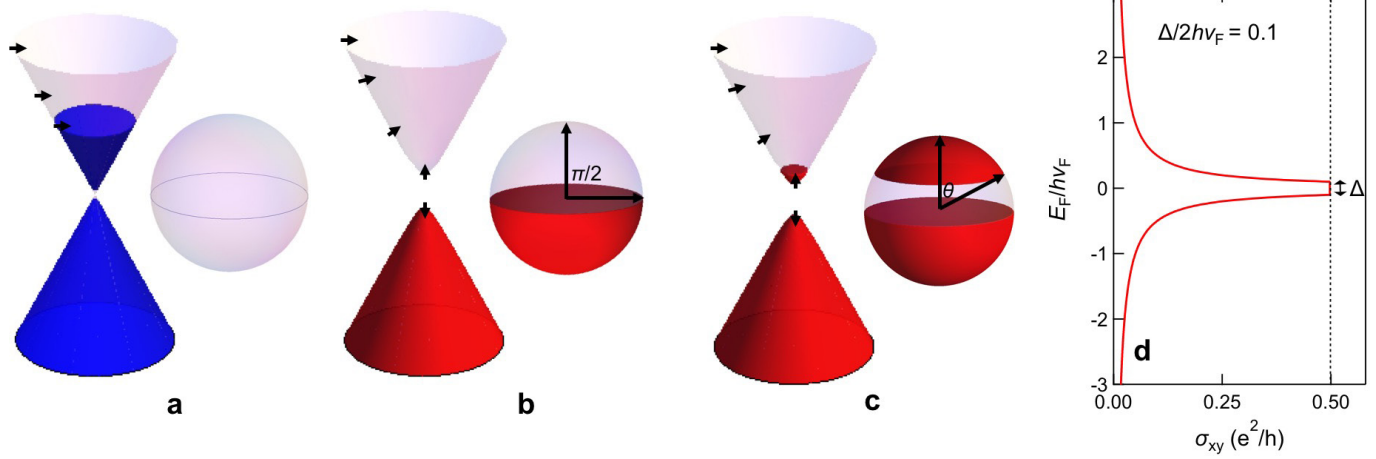
Extended Data Figure 3 | Momentum and energy-dependent band structure along high-symmetry directions. a, e, Fermi surface of Fe_3Sn_2 obtained from different experimental geometries. b–d, f, g, Band dispersion of Fe_3Sn_2 along high-symmetry directions. The panels correspond to the momentum directions along the red (b), orange (c), green (d), magenta (f) and purple (g) dotted lines in a and e. The inset

in d shows the raw data of Fig. 3c (with the same energy and momentum range), highlighting the spectral weight distribution near the Dirac points. h, Energy distribution curves at different K points indicated in c, d, f and g. The curves are shifted along the vertical direction for clarity. The inset shows an example of Gaussian fits; the extracted gap size is $\Delta = 30 \pm 5$ meV.



Extended Data Figure 4 | Photon-energy dependence of ARPES spectra.

ARPES intensity plot for Fe₃Sn₂ taken along the Γ -K direction as a function of binding energy k and photon energy.



Extended Data Figure 5 | Berry curvature and Hall conductivity for a massive Dirac fermion. a–c, Schematics of 2D Dirac fermions and the corresponding Bloch-sphere representation of the wavefunction of filled

states for the gapless case (a) and the gapped case with E_F in (b) and out of (c) the gap. **d,** Fermi energy E_F dependence of σ_{xy} for the case of a single massive Dirac fermion with gap Δ and Fermi velocity v_F .

Timing of oceans on Mars from shoreline deformation

Robert I. Citron^{1,2}, Michael Manga^{1,2} & Douglas J. Hemingway^{1,2}

Widespread evidence points to the existence of an ancient Martian ocean^{1–8}. Most compelling are the putative ancient shorelines in the northern plains^{2,7}. However, these shorelines fail to follow an equipotential surface, and this has been used to challenge the notion that they formed via an early ocean⁹ and hence to question the existence of such an ocean. The shorelines' deviation from a constant elevation can be explained by true polar wander occurring after the formation of Tharsis¹⁰, a volcanic province that dominates the gravity and topography of Mars. However, surface loading from the oceans can drive polar wander only if Tharsis formed far from the equator¹⁰, and most evidence indicates that Tharsis formed near the equator^{11–15}, meaning that there is no current explanation for the shorelines' deviation from an equipotential that is consistent with our geophysical understanding of Mars. Here we show that variations in shoreline topography can be explained by deformation caused by the emplacement of Tharsis. We find that the shorelines must have formed before and during the emplacement of Tharsis, instead of afterwards, as previously assumed. Our results imply that oceans on Mars formed early, concurrent with the valley networks¹⁵, and point to a close relationship between the evolution of oceans on Mars and the initiation and decline of Tharsis volcanism, with broad implications for the geology, hydrological cycle and climate of early Mars.

Distinct geological boundaries (contacts) lining the northern plains of Mars for thousands of kilometres have been interpreted as palaeo-shorelines and evidence of an early ocean^{2–4,6,7}. However, observed long-wavelength deviations (by up to several kilometres) in shoreline elevation from an equipotential have been used as an argument against the emplacement of the contacts by a body of liquid water, the interpretation of the features as shorelines, and the existence of a Martian ocean⁹. Perron *et al.*¹⁰ showed that the elevation changes of two extensive contacts, Arabia (contact 1) and Deuteronilus (contact 2), can be explained by deformation due to 30°–60° and 5°–25° of post-Tharsis true polar wander (TPW), respectively, because a varying rotation pole also changes the orientation of a planet's equatorial bulge, or polar flattening, altering equipotential surfaces (such as sea levels) globally. Such large magnitudes of TPW can be driven by ocean loading/unloading, but only if Tharsis formed far from the equator¹⁰. If Tharsis formed near the equator, then the remnant fossil bulge would have prevented ocean loading from causing large amounts of post-Tharsis TPW (see Extended Data Fig. 1).

Most evidence points to the formation of Tharsis near the equator^{11–15}. Mars' remnant rotational figure (fossil bulge) is close to the equator, indicating a palaeopole of (259.5 ± 49.5° E, 71.1^{+17.5°}_{-14.4°} N), the likely pre-Tharsis orientation of Mars¹⁴. The pre-Tharsis palaeopole also matches the likely orientation of Mars during valley network formation¹⁵. Formation of Tharsis probably drove only limited (approximately 20°) TPW to reach Mars' current configuration, which precludes the possibility that surface loads drove sufficient TPW to deform the shorelines^{10,16}.

We propose that the Arabia shoreline instead formed before or during the early stages of Tharsis emplacement, which initiated > 3.7 billion years (Gyr) ago¹⁷ when the rotation pole of Mars was at the palaeopole (259.5° E, 71.1° N) corresponding to the fossil bulge¹⁴. The Arabia shoreline, potentially emplaced at least 4 Gyr ago⁶, would have been modified by both topographic changes from Tharsis (which dominates Mars' topography and gravity on a global scale; see Extended Data Fig. 2), and the approximately 20° of Tharsis-induced TPW. The Deuteronilus shoreline, which differs less from a present-day equipotential than the older Arabia shoreline, is dated to about 3.6 Gyr ago¹⁸, after most of Tharsis was emplaced. However, Tharsis had complex and multi-stage growth that extended into the Hesperian and Amazonian^{17,19}, meaning that the Deuteronilus shoreline would have been deformed by the late stages of Tharsis' emplacement. We examine a chronology in which shoreline deformation is due mainly to Tharsis (Table 1), and compare expected deformation due to Tharsis with the elevation profiles of the Arabia and Deuteronilus contacts.

Assuming the Arabia shoreline formed before Tharsis, and the Deuteronilus shoreline formed after most of Tharsis was emplaced, we compare the best fits for the deformation expected from Tharsis to the current topography of the shorelines, including an offset factor *Z* to represent sea level at the time of shoreline formation. We also examine the Isidis shoreline, which formed 100 million years (Myr) after Deuteronilus¹⁸. For the Arabia shoreline emplaced before Tharsis, deformation is expressed as the contribution of Tharsis to Mars' topography along the shoreline, and the change in topography from limited Tharsis-induced TPW. For the Deuteronilus and Isidis shorelines emplaced during the late stages of Tharsis growth, deformation is taken as the percentage of Tharsis' contribution to topography occurring after the shorelines formed, and no contribution from TPW (because reorientation should occur within tens of thousands of years to a few million years after the Tharsis plume reaches the surface²⁰, much less than the 100 Myr or more that lies between Tharsis initiation and Deuteronilus formation). See Methods for more details.

We show that the Arabia shoreline's deviations from an equipotential can be explained almost entirely by deformation due to Tharsis emplacement (Fig. 1). Our best fit (equation (3) with *Z* = −2.3 km) yields a root-mean square misfit σ_{rms} of 0.615 km, comparable to the error values from Perron *et al.*¹⁰, and follows the slope of the shoreline data better from 1,000 km to 6,600 km. The limited Tharsis-induced TPW has a negligible effect. A slightly lower σ_{rms} is obtained if only 80% of Tharsis topography was emplaced after the Arabia shoreline formed (Extended Data Fig. 3). However, the difference between the fits using 80% or 100% of Tharsis' topography is negligible considering the scatter in the shoreline data. Our model therefore suggests that the Arabia shoreline formed before or during the early stages of Tharsis' growth.

The Deuteronilus shoreline's deviations from an equipotential can be explained by deformation due to the emplacement of about 17% of Tharsis topography (Fig. 2), indicating that the shoreline formed during the late stages of Tharsis' growth. Our best fit (equation (4) with

¹Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, California, USA. ²Center for Integrative Planetary Science, University of California, Berkeley, Berkeley, California, USA.

Table 1 | Possible evolution of Martian shorelines

Time (epoch)	Event	Effect
At least 4 Gyr ago (Early Noachian)	Arabia shoreline forms	
>3.6 Gyr ago (Late Noachian/Early Hesperian)	Majority of Tharsis forms	Both Tharsis emplacement/loading and Tharsis-induced TPW (20°) deform the Arabia shoreline
3.6 Gyr ago (Early Hesperian)	Deuteronilus shoreline forms	
3.5 Gyr ago (Late Hesperian)	Isidis shoreline forms	
3.6 Gyr ago to present (Early Hesperian–Late Amazonian)	Remainder of Tharsis forms	Remaining Tharsis growth deforms the Arabia, Deuteronilus and Isidis shorelines
3.5–3.0 Gyr ago (Late Hesperian/Early Amazonian)	Loading of Utopia basin	Loading of the Utopia basin tilts the Isidis basin and deforms sections of the Deuteronilus shoreline

The table outlines the sequence of events required to produce a pre- or early-Tharsis Arabia shoreline, and the formation of the Deuteronilus shoreline concurrent with the later stages of Tharsis' growth. Explaining the Isidis shoreline's topography requires subsequent loading of the Utopia basin (see Extended Data Fig. 4 and Methods).

$C=0.17$ and $Z=-3.68$ km) yields $\sigma_{rms}=0.110$ km. Our fit successfully recovers the low elevation of the Phlegra contact, and also captures the decrease in elevation across Utopia and Elysium West. Neither our model nor the Perron *et al.*¹⁰ model captures the full elevation increase of the Tantalus segment, which may result from the topographic bulge from nearby Alba Patera¹⁸. For the Isidis shoreline, subsequent loading of the Utopia basin is also required to explain the shoreline's topography (see Extended Data Fig. 4 and Methods).

The relation between the shorelines and global deformation due to Tharsis and its associated TPW is illustrated in Fig. 3a–c (also see Extended Data Fig. 2). We estimate the volume of water that filled the northern plains to the Deuteronilus and Arabia shorelines by subtracting the relevant Tharsis and TPW contributions from Mars' topography (0.25° per pixel gridded MOLA data²¹) and filling the lowlands to shoreline elevation (Fig. 3d–f). We estimate a Deuteronilus ocean volume of about 1.2×10^7 km³, and an Arabia ocean volume of about 4.1×10^7 km³. These are lower limits because we do not remove excess terrain, such as Elysium, polar deposits, lava/sediment basin deposits, and short-wavelength Tharsis topography (that is, variations in Tharsis topography that occur over short length scales). For the Arabia ocean, use of a map of Mars with excess terrain removed¹⁵ yields an ocean

volume of about 5.5×10^7 km³. The ocean volumes we compute are slightly lower than previous estimates²² because the Tharsis topography we subtract is negative in much of the area enclosed by the northern ocean basin.

Short-wavelength deviations in shoreline elevation from our model may be due to our assumptions that both lithospheric thickness and the rate of Tharsis emplacement are spatially uniform. Spatial variations in lithospheric thickness²³ would allow for non-uniform responses to phenomena such as TPW¹⁰, plate flexure and dynamic topography from the Tharsis plume²⁴. Spatially variable Tharsis emplacement could also affect shoreline modification. Another consideration is ocean loading, but the computed effect on shoreline elevations is small (see Extended Data Fig. 5 and Methods).

Several other short- and long-wavelength processes could have deformed the shorelines in the 3.5 Gyr or more since their emplacement, including dynamic topography from mantle plumes²⁴, lithospheric deformation^{22,25}, glacial erosion²⁶ and plate flexure from loading/unloading. For example, loading of the Utopia basin may have tilted Isidis (see Methods) and deformed sections of the Deuteronilus shoreline¹⁸. Other loads that post-date shoreline formation include Elysium, the polar deposits, and sections of Tharsis. Such loads

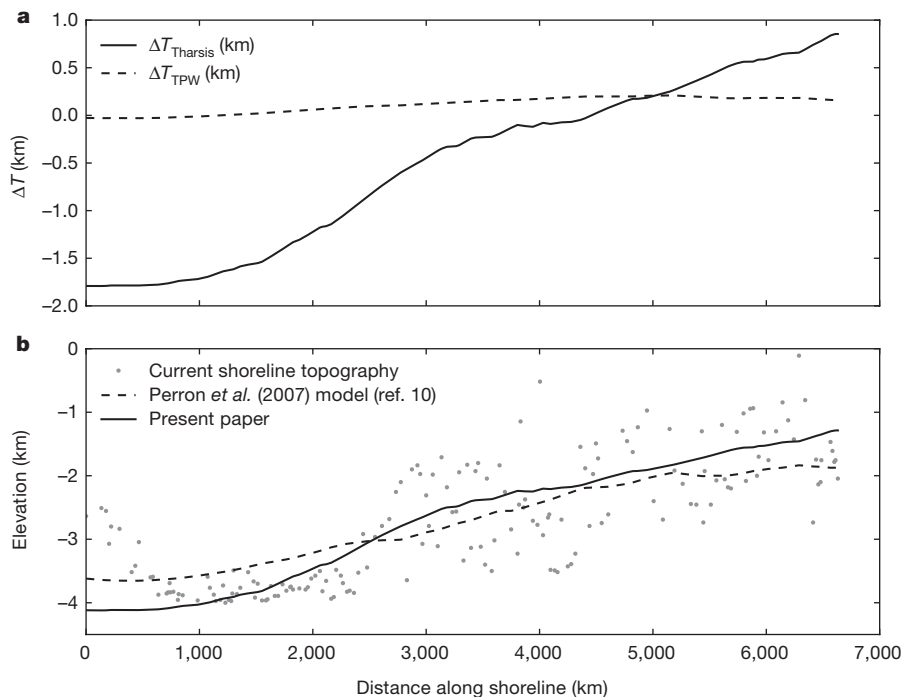


Figure 1 | Comparison of Arabia shoreline topography to shoreline deformation models. a, Change in topography ΔT caused by TPW of 20° (equation (1)) and Tharsis uplift (equation (2)), illustrating that the latter is much more important. **b**, Current topography of the Arabia shoreline from Perron *et al.*¹⁰ (data originally from ref. 7) compared to the Perron

*et al.*¹⁰ model of deformation due to post-Tharsis TPW (with $T_e=200$ km) and our model of deformation due to Tharsis emplacement and induced TPW ($\Delta T_{Tharsis} + \Delta T_{TPW} = 2.3$ km). The starting point for the shoreline is (24.91° W, 13.48° N).

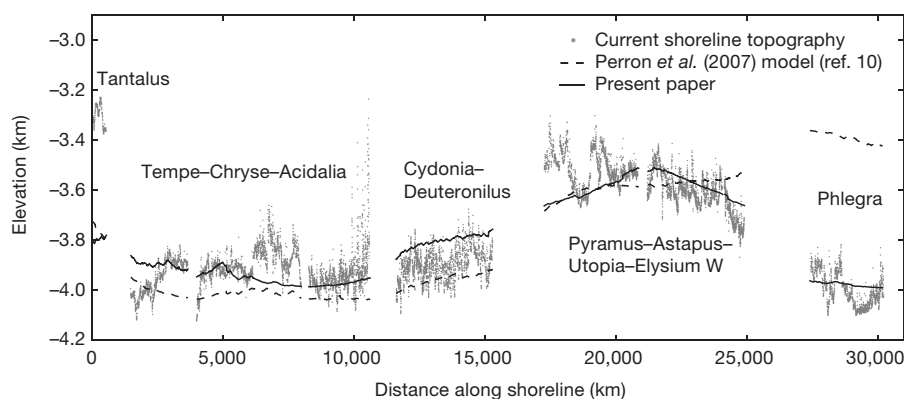


Figure 2 | Comparison of Deuteronilus shoreline topography to shoreline deformation models. Current Deuteronilus topography (data and contact names from ref. 18) compared to the Perron *et al.*¹⁰ model

and our model of deformation due to partial Tharsis emplacement ($0.17\Delta T_{\text{Tharsis}} - 3.68$ km). The starting point for the shoreline is (96.40° W, 63.69° N).

could also induce a small amount ($<2^\circ$) of post-Tharsis TPW¹⁶. Plate flexure associated with impact basins could also deform shorelines. While basins of over 1,000 km in diameter pre-date the Deuteronilus shoreline, some basins may have been coincident with or post-date the Arabia shoreline. Short-wavelength deformation may also be a consequence of the difficulty in identifying the shorelines themselves¹⁸.

Increased accuracy in shoreline identification and dating¹⁸ can help to reconstruct the history of shoreline formation and modification. Several potential shorelines⁶, such as the Ismenius, Acidalia, Elysium, Aeolis and Meridiani contacts, have been relatively unexamined owing to their high degree of discontinuity⁷. Shorelines may also be mismapped; for example, portions of the Meridiani shoreline may be part of the Arabia shoreline²². A re-evaluation of shorelines with full consideration of the various deformation processes may enable the development of a chronology of oceans on Mars. In particular, the Meridiani shoreline^{6,27} may pre-date the Arabia shoreline and have contained a larger volume of water²².

Accurate dating of the Arabia shoreline is necessary to determine whether the shoreline formed before or during the early stages of Tharsis' growth. Formation of the Arabia shoreline after some limited

early Tharsis growth is suggested by Arabia segments that border Acheron Fossae and Tempe Terra⁶, two of the oldest Tharsis units, which are located well north of the expected pre-Tharsis crustal dichotomy boundary (the stark difference in elevation and crustal thickness between the northern lowlands and southern highlands). However, it is possible that the Acheron Fossae and Tempe Terra contacts were misidentified as belonging to the Arabia shoreline, or that the Arabia shoreline initially followed the pre-Tharsis dichotomy boundary, and formed the Tempe Terra and Acheron Fossae contacts only after early Tharsis uplift and deposition.

The decline in ocean volume from the pre- or early-Tharsis Arabia shoreline to the late-Tharsis Deuteronilus shoreline suggests that Tharsis volcanism may have played a critical part in the evolution of a Martian ocean. After Tharsis was mostly emplaced, by about 3.6 Gyr ago, only short-lived lakes may have been stable²⁸, although a Late Hesperian/Early Amazonian ocean has also been suggested on the basis of tsunami evidence^{8,29}. Outgassing from Tharsis could have contributed to either heating³⁰ or cooling³¹ the planet, both of which could produce the decrease in ocean volume from the Arabia shoreline to the Deuteronilus shoreline as Tharsis activity declined. Either a

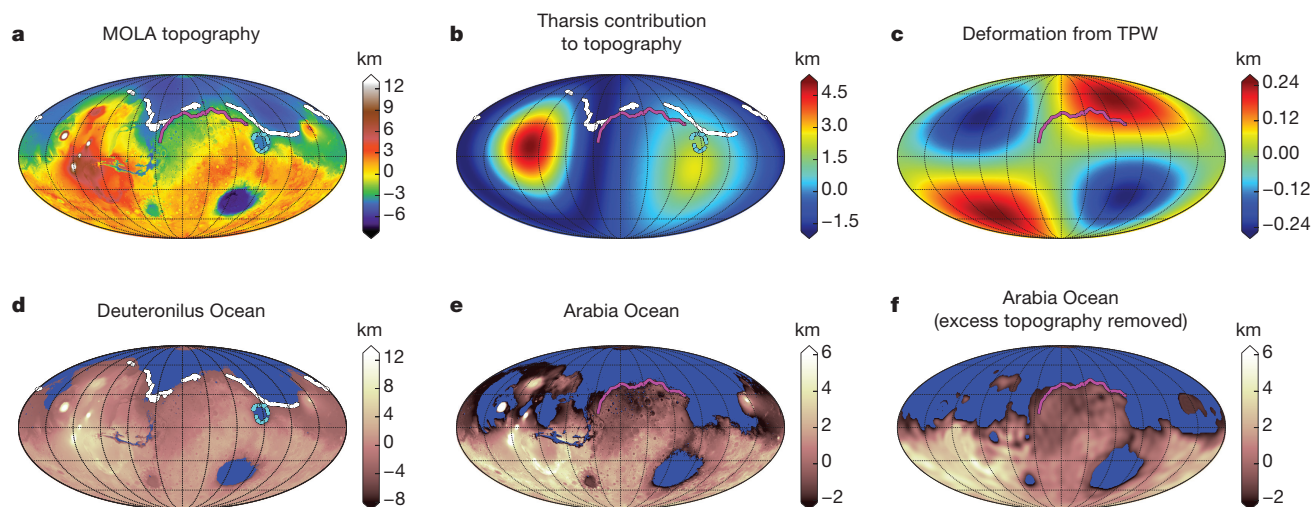


Figure 3 | Shoreline locations relative to current topography, deformation due to Tharsis/TPW, and computed ocean extents. **a**, MOLA topography. **b**, Tharsis contribution to topography (equation (2)). **c**, Deformation due to approximately 20° TPW from the palaeopole corresponding to the fossil bulge¹⁴ with $T_c = 58$ km (equation (1)). **d**, Ocean basin filled to the Deuteronilus shoreline, with the topography of Mars at the time of the Deuteronilus shoreline's formation (MOLA topography minus 17% of Tharsis topography). **e**, Ocean basin filled to the Arabia shoreline, with the topography of Mars at the time of the Arabia

shoreline's formation (MOLA topography minus Tharsis topography and deformation due to TPW). Short-wavelength remnants of Tharsis are visible because Tharsis topography is only modelled up to degree-5. **f**, The 'Mars without Tharsis' map¹⁵, which is similar to **e**, but with short-wavelength Tharsis, Elysium, and polar topography removed, filled to the Arabia shoreline. Shorelines are plotted for Deuteronilus (white), Arabia (magenta) and Isidis (cyan), and colour scales denote elevation (or changes in elevation) in kilometres.

large ocean was in place before Tharsis volcanism initiated, and shrank as Tharsis volcanism cooled the planet, or an ocean arose as a result of heating caused by Tharsis outgassing and decreased in volume as Tharsis volcanism declined. It is also possible that each shoreline represents the transient warming of an otherwise frozen ocean or glacial state³⁰, producing a liquid ocean in periods of heightened Tharsis activity (which, owing to enhanced surface heat flux, may also have resulted in catastrophic circum-Tharsis groundwater discharge¹). If episodic warming was sufficient to melt most of Mars' glaciers, the decrease in ocean volume may record a declining surface water budget. Although geochemical evidence for a northern ocean is ambiguous³², an ocean supported by the degassing of sulfur from Tharsis could explain the lack of widespread carbonate deposits observed in the northern plains³³.

The evolution of water on Mars is critical to understanding the past climate and habitability of the planet. Although shorelines on Mars have provided compelling evidence for a Martian ocean, to explain their deviations from an equipotential has been a challenge. We show that the topography of Martian shorelines can be quantitatively explained by deformation due to the emplacement of Tharsis and resulting TPW (in the case of the Arabia shoreline) or by the latter stages of Tharsis emplacement (in the case of the Deuteronilus shoreline). Formation of the Arabia shoreline before (or during the early stages of) Tharsis emplacement suggests that the Arabia ocean was concurrent with valley network incision¹⁵, which probably occurred as part of a globally active hydrosphere capable of supporting such an ocean⁵. The consistency between the topography of the Martian shorelines, their ages, and the chronology of topographic changes due to Tharsis emplacement and associated TPW, suggests that the Arabia and Deuteronilus contacts are evidence that Martian oceans existed, and may have been linked to Tharsis volcanism.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2017; accepted 25 January 2018.

Published online 19 March 2018.

- Baker, V. R. *et al.* Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature* **352**, 589–594 (1991).
- Parker, T. J., Saunders, R. S. & Schneeberger, D. M. Transitional morphology in West Deuteronilus Mensae, Mars: implications for modification of the lowland/upland boundary. *Icarus* **82**, 111–145 (1989).
- Parker, T. J., Gorsline, D. S., Saunders, R. S., Pieri, D. C. & Schneeberger, D. M. Coastal geomorphology of the Martian northern plains. *J. Geophys. Res. Planets* **98**, 11061–11078 (1993).
- Head, J. W. *et al.* Possible ancient oceans on Mars: evidence from Mars Orbiter Laser Altimeter data. *Science* **286**, 2134–2137 (1999).
- Di Achille, G. & Hynek, B. M. Ancient ocean on Mars supported by global distribution of deltas and valleys. *Nat. Geosci.* **3**, 459–463 (2010).
- Clifford, S. M. & Parker, T. J. The evolution of the Martian hydrosphere: implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* **154**, 40–79 (2001).
- Carr, M. H. & Head, J. W. Oceans on Mars: an assessment of the observational evidence and possible fate. *J. Geophys. Res. Planets* **108**, 5042 (2003).
- Rodríguez, J. A. P. *et al.* Tsunami waves extensively resurfaced the shorelines of an early Martian ocean. *Sci. Rep.* **6**, 25106 (2016).
- Malin, M. C. & Edgett, K. S. Oceans or seas in the Martian northern lowlands: high resolution imaging tests of proposed coastlines. *Geophys. Res. Lett.* **26**, 3049–3052 (1999).
- Perron, J. T., Mitrovica, J. X., Manga, M., Matsuyama, I. & Richards, M. A. Evidence for an ancient martian ocean in the topography of deformed shorelines. *Nature* **447**, 840–843 (2007).
- Willemann, R. J. Reorientation of planets with elastic lithospheres. *Icarus* **60**, 701–709 (1984).
- Roberts, J. H. & Zhong, S. The cause for the north-south orientation of the crustal dichotomy and the equatorial location of Tharsis on Mars. *Icarus* **190**, 24–31 (2007).
- Daradich, A. *et al.* Equilibrium rotational stability and figure of Mars. *Icarus* **194**, 463–475 (2008).
- Matsuyama, I. & Manga, M. Mars without the equilibrium rotational figure, Tharsis, and the remnant rotational figure. *J. Geophys. Res. Planets* **115**, E12020 (2010).
- Bouley, S. *et al.* Late Tharsis formation and implications for early Mars. *Nature* **531**, 344–347 (2016).
- Kite, E. S., Matsuyama, I., Manga, M., Perron, J. T. & Mitrovica, J. X. True Polar wander driven by late-stage volcanism and the distribution of paleopolar deposits on Mars. *Earth Planet. Sci. Lett.* **280**, 254–267 (2009).
- Anderson, R. C. *et al.* Primary centers and secondary concentrations of tectonic activity through time in the western hemisphere of Mars. *J. Geophys. Res. Planets* **106**, 20563–20585 (2001).
- Ivanov, M. A., Erkeling, G., Hiesinger, H., Bernhardt, H. & Reiss, D. Topography of the Deuteronilus contact on Mars: evidence for an ancient water/mud ocean and long-wavelength topographic readjustments. *Planet. Space Sci.* **144**, 49–70 (2017).
- Dohm, J. M., Baker, V. R., Maruyama, S. & Anderson, R. C. In *Superplumes: Beyond Plate Tectonics* 523–536 (Springer, 2007).
- Rouby, H., Greff-Lefftz, M. & Besse, J. Rotational bulge and one plume convection pattern: influence on Martian true polar wander. *Earth Planet. Sci. Lett.* **272**, 212–220 (2008).
- Smith, D. E., Zuber, M. T., Neumann, G. A., Guinness, E. A. & Slaveny, S. Mars Global Surveyor Laser Altimeter mission experiment gridded data record. MGS-M-MOLA-5-MEGDR-L3-V1.0. <http://pds-geosciences.wustl.edu/missions/mgs/megdr.html> (NASA Planetary Data System, 2003).
- Ruiz, J., Tejero, R., Gómez-Ortiz, D. & López, V. in *Space Science: New Research* 141–164 (Nova Science, 2006).
- Grott, M. & Breuer, D. On the spatial variability of the Martian elastic lithosphere thickness: evidence for mantle plumes? *J. Geophys. Res. Planets* **115**, E03005 (2010).
- Roberts, J. H. & Zhong, S. Plume-induced topography and geoid anomalies and their implications for the Tharsis rise on Mars. *J. Geophys. Res. Planets* **109**, E03009 (2004).
- Ruiz, J., Fairén, A. G., Dohm, J. M. & Tejero, R. Thermal isostasy and deformation of possible paleoshorelines on Mars. *Planet. Space Sci.* **52**, 1297–1301 (2004).
- Davila, A. F. *et al.* Evidence for Hesperian glaciation along the Martian dichotomy boundary. *Geology* **41**, 755–758 (2013).
- Edgett, K. S. & Parker, T. J. Water on early Mars: possible subaqueous sedimentary deposits covering ancient cratered terrain in western Arabia and Sinus Meridiani. *Geophys. Res. Lett.* **24**, 2897 (1997).
- Kite, E. S. *et al.* Methane bursts as a trigger for intermittent lake-forming climates on post-Noachian Mars. *Nat. Geosci.* **10**, 737–740 (2017).
- Costard, F. *et al.* Modeling tsunami propagation and the emplacement of thumbprint terrain in an early Mars ocean. *J. Geophys. Res. Planets* **122**, 633–649 (2017).
- Halevy, I. & Head, J. W. Episodic warming of early Mars by punctuated volcanism. *Nat. Geosci.* **7**, 865–868 (2014).
- Tian, F. *et al.* Photochemical and climate consequences of sulfur outgassing on early Mars. *Earth Planet. Sci. Lett.* **295**, 412–418 (2010).
- Pan, L., Ehlmann, B. L., Carter, J. & Ernst, C. M. The stratigraphy and history of Mars' northern lowlands through mineralogy of impact craters: a comprehensive survey. *J. Geophys. Res. Planets* **122**, 1824–1854 (2017).
- Halevy, I., Zuber, M. T. & Schrag, D. P. A sulfur dioxide climate feedback on early Mars. *Science* **318**, 1903–1907 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. T. Perron for providing the data for the Arabia shoreline (originally from ref. 7), and M. A. Ivanov for providing the data for the Deuteronilus and Isidis shorelines. We thank I. Matsuyama for discussions regarding this research. R.I.C. and M.M. are supported by NSF EAR-1135382. D.J.H. is supported by the Miller Institute for Basic Research in Science.

Author Contributions All authors discussed the research idea, methods and interpretation of results. R.I.C. and M.M. developed the hypothesis with input from D.J.H. R.I.C. performed the calculations and wrote the manuscript, with guidance, comments, and revisions from M.M. and D.J.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.I.C. (ricitron@berkeley.edu).

Reviewer Information *Nature* thanks S. Bouley and M. Zuber for their contribution to the peer review of this work.

METHODS

Arabia shoreline (pre- or early-Tharsis formation). We assume deformation of the Arabia shoreline since its formation is due to global changes in topography resulting from Tharsis' formation (emplacement and loading) and the approximately 20° of Tharsis-induced TPW (Table 1).

The topographic response to TPW is given by the change in the flattening of the planet caused by the difference between the centrifugal potential at the initial and final rotation poles¹⁰. For a shoreline in place before TPW occurs, the deformation of the shoreline topography due to TPW¹⁰ is:

$$\Delta T_{\text{TPW}}(\theta, \varphi) = \frac{\omega^2 a^2}{3g} [P_{2,0}(\cos\gamma) - P_{2,0}(\cos\theta)] [h_2 - (1 + k_2)] \quad (1)$$

where a is the mean planetary radius, ω is the rotation rate, g is the surface gravity, γ is the angular distance between a given current colatitude and longitude (θ, φ) and the palaeopole and h_2 and k_2 are the secular (fluid-limit) degree-2 Love numbers that depend on the density and elastic structure of Mars. The unnormalized degree-2 Legendre polynomial $P_{2,0}(\cos\eta) = \frac{1}{2}(3\cos^2\eta - 1)$.

The change in topography due to the emplacement of Tharsis and its associated loading is:

$$\Delta T_{\text{Tharsis}}(\theta, \varphi) = S_{\text{Tharsis}}(\theta, \varphi) - N_{\text{Tharsis}}(\theta, \varphi) \quad (2)$$

where S_{Tharsis} and N_{Tharsis} are Tharsis' contribution to the shape and geoid of Mars, respectively. We use gravity and shape coefficients for Tharsis up to degree-5 from Matsuyama and Manga¹⁴.

The current topography of the Arabia shoreline should therefore follow the deformation profile ΔT_A , given by:

$$\Delta T_A(\theta, \varphi) = \Delta T_{\text{TPW}}(\theta, \varphi) + \Delta T_{\text{Tharsis}}(\theta, \varphi) + Z \quad (3)$$

where Z is a constant to adjust for the sea level at the time of the shoreline's emplacement. We minimize the least-squares misfit (σ_{rms}) between equation (3) and the shoreline elevation data for the Arabia contact examined in ref. 10 (data originally from ref. 7). We assume a fixed palaeopole (259.5° E, 71.1° N), corresponding to the fossil bulge¹⁴. We use an elastic lithosphere thickness $T_e = 58$ km, the expected value at the time of Tharsis' emplacement¹⁴, corresponding to $h_2 = 2.0$ and $k_2 = 1.1$ ¹⁵.

We also test whether the Arabia shoreline can be explained by deformation due to only a certain percentage of Tharsis' emplacement and associated loading, by multiplying $\Delta T_{\text{Tharsis}}$ in equation (3) by a factor C , corresponding to the percentage of Tharsis topography emplaced after shoreline formation (see Extended Data Fig. 3).

Deuteronilus shoreline (late-stage Tharsis formation). The Deuteronilus shoreline post-dates the initiation of Tharsis by >100 Myr, and therefore probably formed after a large portion of Tharsis was emplaced. The shoreline also probably post-dates most Tharsis-induced TPW, which should have occurred within a few million years of load emplacement²⁰. Estimates of load-driven TPW on Mars suggest timescales less than 10 Myr^{34–36}, well within the required pre-Deuteronilus timescale. Although a fraction of the 20° of Tharsis-induced TPW may be due to relaxation of the lithosphere and occur on longer timescales³⁷, this should have a negligible effect given the small influence of TPW on shoreline deformation (Fig. 1a). Accordingly, we assume that deformation to the Deuteronilus shoreline since its formation is due to the topographic response of Mars to only the late stages of Tharsis' emplacement and associated loading. The current topography of the Deuteronilus shoreline should therefore follow the deformation profile ΔT_D , given by:

$$\Delta T_D(\theta, \varphi) = C \Delta T_{\text{Tharsis}}(\theta, \varphi) + Z \quad (4)$$

where $\Delta T_{\text{Tharsis}}$ is deformation due to Tharsis (equation (2)), C is a constant representing how much of Tharsis formed after Deuteronilus was formed, and Z is a constant to adjust for sea level at the time of the shoreline's formation. We minimize the misfit between equation (4) and the Deuteronilus shoreline elevation data¹⁸, to determine the optimal amount of Tharsis topography that should post-date the shoreline's formation.

Isidis shoreline (late-stage Tharsis formation). Because the Isidis shoreline is 100 Myr younger than Deuteronilus, we assume a similar deformation profile and compare equation (4) to the Isidis shoreline data¹⁸. We use the same value of C optimized for the Deuteronilus shoreline, but allow Z to vary, reflecting that sea level could change in the 100 Myr between the formation of the Deuteronilus contact and the Isidis contact, but deformation from Tharsis topography should not change substantially. Although C should be slightly less for Isidis, optimizing for C would result in an unrealistic $C = 0$ because deformation due to Tharsis along the Isidis shoreline results in a tilt that is opposite to the present tilt (Extended Data Fig. 4).

For the Isidis shoreline, our model predicts that deformation due to Tharsis would have tilted Isidis opposite to its present tilt (Extended Data Fig. 4). While this appears contradictory, the mismatch is possible if Isidis was tilted to its present orientation by loading of the Utopia basin. The Utopia basin has a large positive gravity anomaly^{38,39}, indicating about 18 km of excess fill⁴⁰. Such a load would have caused elastic plate flexure and a peripheral bulge, which could have tilted the Isidis basin. Using a plate flexure model, McGowan and McGill⁴¹ show that loading of Utopia could have tilted Isidis to an even greater extent than currently observed. Therefore, some amount of reverse tilting (as our model predicts) is possible. The timing of Utopia loading relative to the subsequent Tharsis deformation is irrelevant provided that Utopia loading also occurred after the Isidis shoreline formed. We expect loading of Utopia to occur after Isidis shoreline formation because a shrinking Martian ocean would evaporate from the Utopia basin last, depositing the non-volatile component of the ocean there. Additionally, if the ocean became cold and glacial during its decline⁴², then receding glaciers may also have loaded Utopia with excess sediment. The deposits in the base of Utopia basin date to the early Amazonian (<3–3.46 Gyr ago)^{43,44}, after the emplacement of the Isidis shoreline. The eastern portion of Utopia also contains volcanic deposits from Elysium that date to the Amazonian⁴⁴, which could also contribute to loading. While loading from the ocean itself is expected to produce some plate flexure, it is not sufficient to explain the tilt of the Isidis basin (see Extended Data Fig. 5c), and water loading/unloading of Utopia is also insufficient to explain Isidis' tilt⁴¹. Therefore, deposition of material from a receding liquid, muddy or frozen ocean may explain the tilt of the Isidis basin, even if some reverse tilting is caused by deformation due to Tharsis.

Effect of elastic lithosphere thickness. The gravity and shape coefficients we use to subtract Tharsis topography are based on an assumed $T_e = 58$ km, the expected value at the time of Tharsis loading¹⁴. However, the estimate of T_e yields a 90% confidence interval with a minimum and maximum of 26 km and 92 km, respectively¹⁴. A thinner or thicker T_e would alter the deformation due to Tharsis (and TPW) because the Love numbers used to compute Mars' deformation would change. To estimate the effect of $T_e = 26$ km or 92 km on deformation due to Tharsis, we recompute Tharsis' gravity and shape coefficients following the method of ref. 14. Using a fixed Tharsis centre location (258.6° E, 9.8° N), Matsuyama and Manga¹⁴ compute the degree-2 gravity coefficients of Tharsis using a minimization technique with four unconstrained model parameters (T_e , non-dimensional Tharsis load Q , palaeopole colatitude θ_R and palaeopole longitude φ_R), where the palaeopole corresponds to the axis of rotation when the fossil (remnant) bulge was formed. This results in probability density functions for each unconstrained parameter, with the weighted averages (expected values) used to compute the gravity and shape coefficients. We redo this analysis, as described in section 5 of ref. 14, but with T_e treated as a constrained parameter. This allows us to estimate the expected values of Q , θ_R and φ_R for a given value of T_e . We find that for $T_e = 26$ km, $\bar{Q} = 3.95$, $\bar{\theta}_R = 17.9^\circ$, and $\bar{\varphi}_R = 259.1^\circ$. For $T_e = 92$ km, $\bar{Q} = 1.57$, $\bar{\theta}_R = 14.2^\circ$, and $\bar{\varphi}_R = 259.3^\circ$. Tharsis' degree-2 gravity coefficients are recomputed using these values. The degree-3 to -5 gravity coefficients of Tharsis are computed from minimization against the observed degree-3 to -5 gravity coefficients, and are therefore not dependent on T_e . Shape coefficients for Tharsis are computed up to degree 5 following section 7 of ref. 14. We compute the load Love numbers using the ALMA code⁴⁵, with a five-layer model as described in ref. 14.

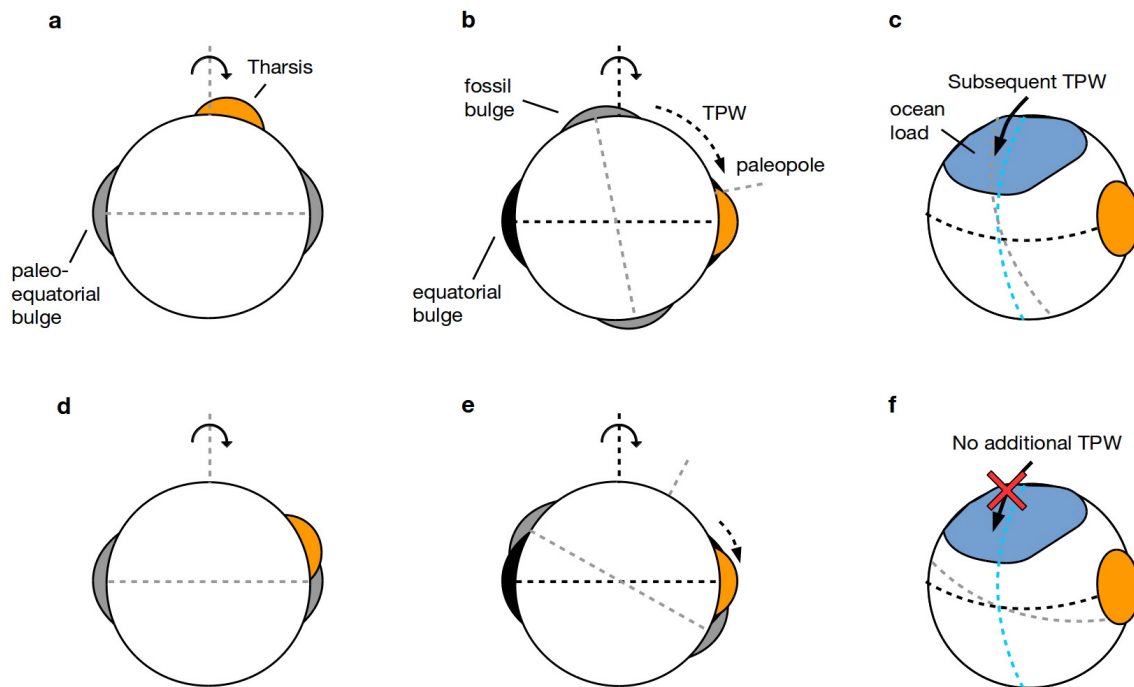
We construct new best-fit deformation profiles for $T_e = 26$ km and 92 km, but with the corresponding Tharsis gravity and shape coefficients that we computed for each T_e . The best-fit profiles for $T_e = 26$ km and 92 km are compared with the nominal $T_e = 58$ km profiles in Extended Data Fig. 3. All best-fit profiles are relatively similar, showing that changes in T_e do not have much effect on our conclusions.

Effect of plate flexure. Although Perron *et al.*¹⁰ found that plate flexure due to loading of the ocean basin should not substantially affect the shoreline elevations, their analysis was for $T_e = 200$ km, whereas we use $T_e = 58$ km. The ocean basin resulting from our analysis also has less volume and a different shape, because we subtract Tharsis topography, which has a negative component in much of the Borealis basin. To compute plate flexure due to ocean loading, we expand the surface density of the ocean load in spherical harmonics and compute the associated displacement using the method described in ref. 46. For the Arabia ocean, the ocean load is computed by subtracting the pre-Tharsis topography of Mars from the best-fit Arabia ocean elevation ($Z = -2.3$ km). The pre-Tharsis Martian topography is computed by subtracting the deformation due to Tharsis and TPW, equations (2) and (1), from Mars' current topography (0.25° per pixel gridded MOLA data²¹). For the ocean level corresponding to the Deuteronilus and Isidis shorelines, only 17% of deformation due to Tharsis was subtracted from Mars' current topography, and the ocean elevation Z was set to -3.68 km and -3.95 km, respectively. We use a Young's modulus of 70 GPa, Poisson ratio of 0.25, and an assumed value of $T_e = 58$ km. The loaded shoreline topography is compared to the

unloaded topography in Extended Data Fig. 5. We compute a maximum magnitude of deflection of 134 m, 84 m and 57 m, for the Arabia, Deuteronilus and Isidis shorelines, respectively. The mean magnitude of deflection is 35 m for the Arabia shoreline and 17 m for the Deuteronilus and Isidis shorelines. Deformation of the shorelines due to unloading of the ocean basin is negligible.

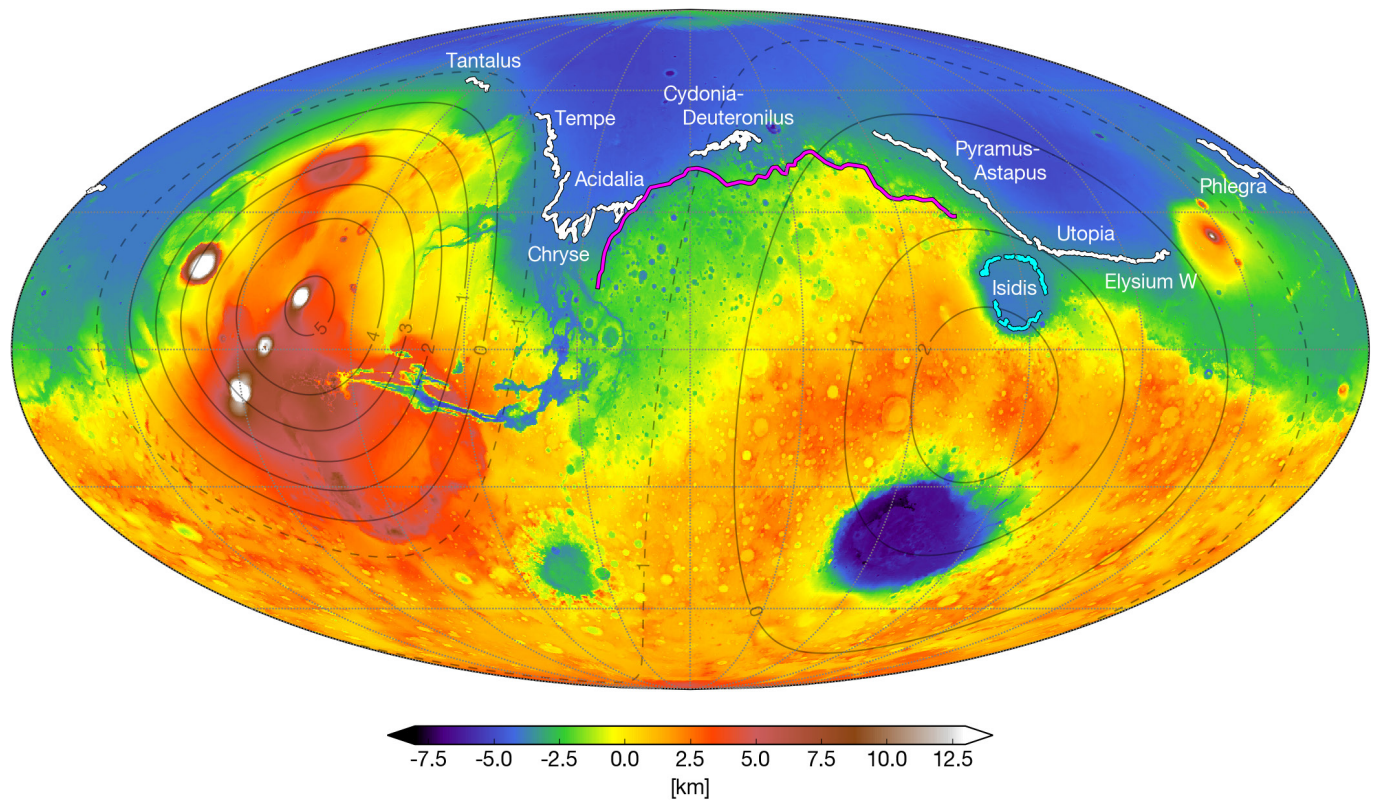
Data availability. The data that supports the findings of this study are available on request from the corresponding author. Gravity and shape coefficients for Tharsis are included in the Supplementary Information. Shoreline data should be requested from the respective sources.

34. Harada, Y. Long-term polar motion on a quasi-fluid planetary body with an elastic lithosphere: semi-analytic solutions of the time-dependent equation. *Icarus* **220**, 449–465 (2012).
35. Chan, N.-H. *et al.* Time-dependent rotational stability of dynamic planets with elastic lithospheres. *J. Geophys. Res. Planets* **119**, 169–188 (2014).
36. Hu, H., van der Wal, W. & Vermeersen, L. L. A. A full-Maxwell approach for large angle polar wander of viscoelastic bodies. *J. Geophys. Res. Planets* **122**, 2745–2764 (2017).
37. Moore, K. M., Chan, N.-H., Daradich, A. & Mitrovica, J. X. Time-dependent rotational stability of dynamic planets with viscoelastic lithospheres. *Icarus* **289**, 34–41 (2017).
38. Sjogren, W. L. Mars gravity: high-resolution results from Viking Orbiter 2. *Science* **203**, 1006–1010 (1979).
39. Zuber, M. T. *et al.* Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* **287**, 1788–1793 (2000).
40. Searls, M. L., Banerdt, W. B. & Phillips, R. J. Utopia and Hellas basins, Mars: twins separated at birth. *J. Geophys. Res. Planets* **111**, E08005 (2006).
41. McGowan, E. M. & McGill, G. E. Anomalous tilt of Isidis Planitia, Mars. *Geophys. Res. Lett.* **33**, L08S06 (2006).
42. Fairén, A. G. A cold and wet Mars. *Icarus* **208**, 165–175 (2010).
43. Werner, S. C. & Tanaka, K. L. Redefinition of the crater-density and absolute-age boundaries for the chronostratigraphic system of Mars. *Icarus* **215**, 603–607 (2011).
44. Tanaka, K. L., Robbins, S. J., Fortezzo, C. M., Skinner, J. A. & Hare, T. M. The digital global geologic map of Mars: chronostratigraphic ages, topographic and crater morphologic characteristics, and updated resurfacing history. *Planet. Space Sci.* **95**, 11–24 (2014).
45. Spada, G. ALMA, a Fortran program for computing the viscoelastic Love numbers of a spherically symmetric planet. *Comput. Geosci.* **34**, 667–687 (2008).
46. Arkani-Hamed, J. The lunar mascons revisited. *J. Geophys. Res. Planets* **103**, 3709–3739 (1998).



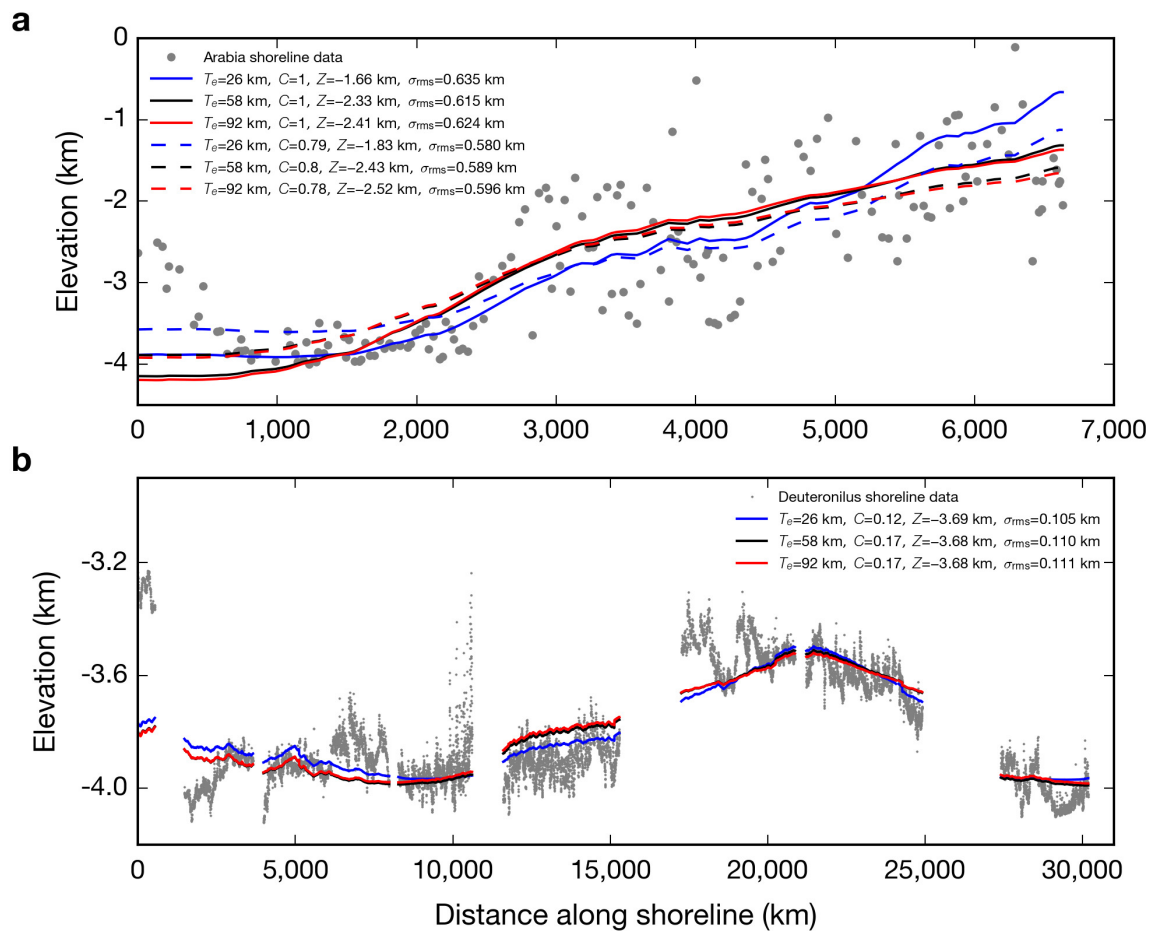
Extended Data Figure 1 | Illustration of the feasibility of post-Tharsis TPW depending on the location of Tharsis' formation. **a–c,** Tharsis (orange zone) forms far from the palaeo-equator (**a**), causing large-scale TPW as the planet reorients so Tharsis is at the equator (**b**). After the reorientation, the fossil bulge is far from the current equatorial bulge, making the rotation pole sufficiently unstable to allow for ocean loading (blue zone) to cause subsequent TPW along an arc 90° from Tharsis

(**c**, blue dashed line)¹⁰. **d–f,** Alternatively, Tharsis forms near the palaeo-equator (**d**), causing limited (approximately 20°) TPW (**e**). The position of the fossil bulge near the equator stabilizes the planet against subsequent TPW caused by oceans and other surface loads (**f**, see supplementary figure 1 of ref. 10). Formation of Tharsis near the equator is supported by refs 14 and 15.



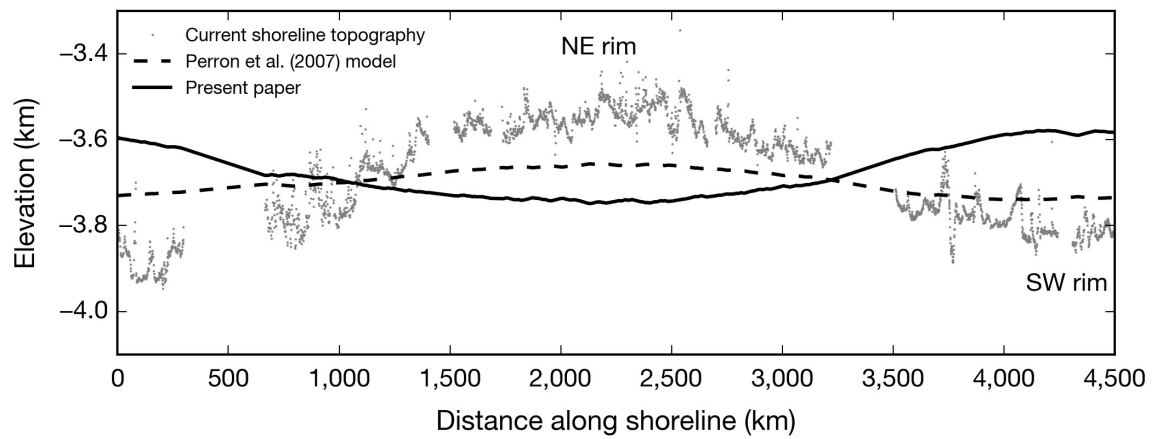
Extended Data Figure 2 | Map of shoreline locations, MOLA topography, and Tharsis deformation. Arabia (magenta) shoreline data are from ref. 10 (data originally from ref. 7). Deuteronilus (white)

and Isidis (cyan) shoreline data and regional names are from ref. 18. The contribution of Tharsis to Mars' topography up to degree-5 (equation (2)) is displayed as 1-km dark grey contours (dashed contours are negative).



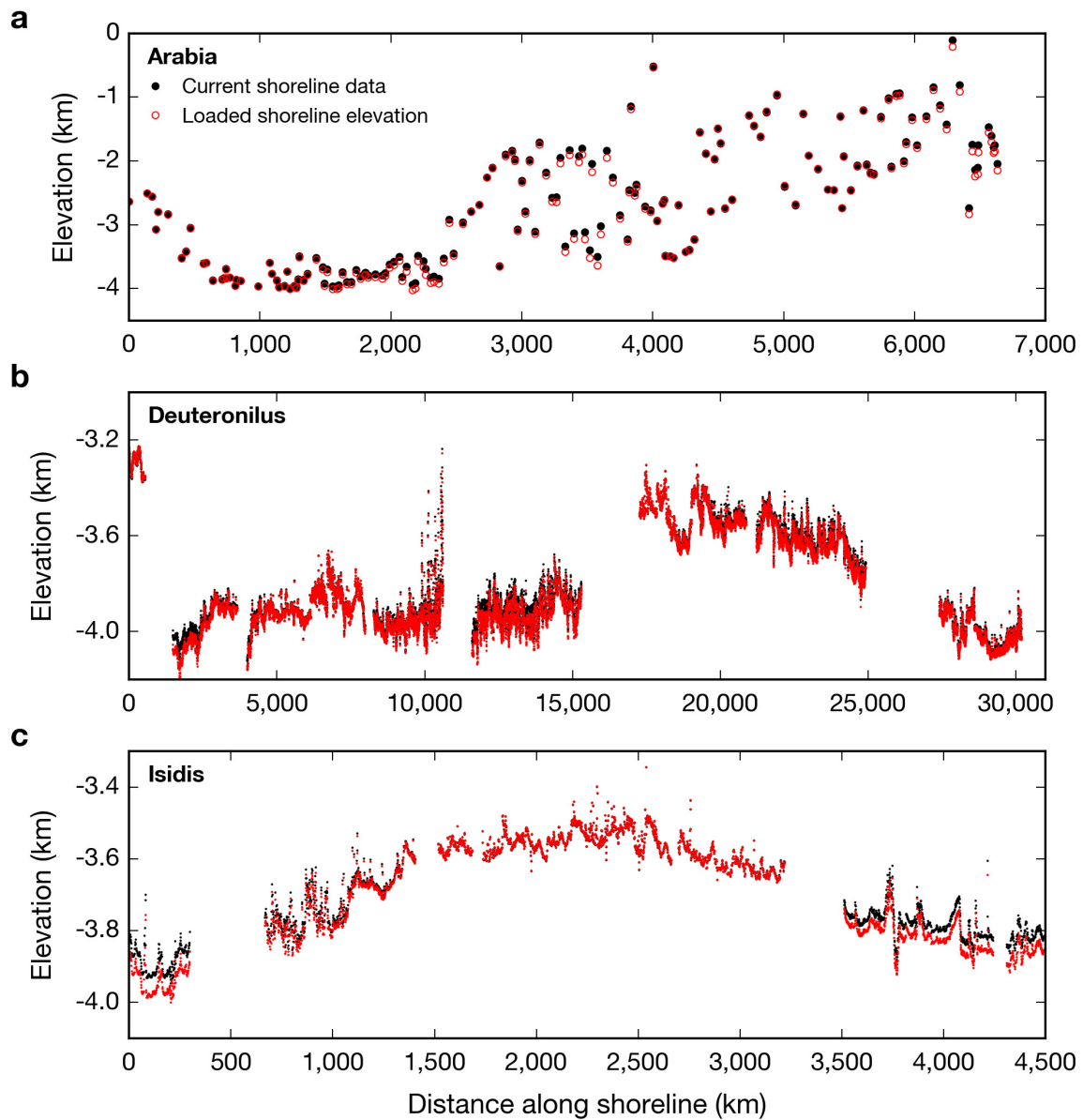
Extended Data Figure 3 | Effect of elastic lithosphere thickness on deformation due to Tharsis. **a**, Current Arabia shoreline topography compared to displacement due to TPW and Tharsis deformation (equation (3)). Tharsis gravity and shape coefficients are computed separately for $T_e = 26$ km, 58 km and 92 km (see Methods), which each yield a corresponding best-fit offset Z and error σ_{rms} . Dashed lines

show the best fit when the percentage of Tharsis topography added after shoreline formation was allowed to vary by a factor C . Solid lines assume 100% percent of Tharsis topography was emplaced after shoreline formation ($C = 1$). **b**, Deuteronilus shoreline topography compared to the best-fit displacement due to Tharsis loading (equation (4)) for $T_e = 26$ km, 58 km and 92 km.



Extended Data Figure 4 | Comparison of Isidis shoreline topography to shoreline deformation models. Current Isidis shoreline topography (elevation data from ref. 18) compared to the Perron *et al.*¹⁰ model for $T_c = 200$ km and our model of deformation due to partial Tharsis emplacement ($0.17\Delta T_{\text{Tharsis}} - 3.95$ km). The topography of the Isidis

shoreline can be explained by subsequent loading of the Utopia basin (see Methods). The starting point for the shoreline is (82.32° E, 7.36° N), near the southwest (SW) rim, with shoreline data proceeding clockwise through the northeast (NE) rim.



Extended Data Figure 5 | Effect of plate flexure due to ocean loading on shoreline topography. Current shoreline elevations are plotted against displaced elevations for the Arabia shoreline (a), the Deuteronilus shoreline (b) and the Isidis shoreline (c).

Insulin resistance in cavefish as an adaptation to a nutrient-limited environment

Misty R. Riddle^{1*}, Ariel C. Aspiras^{1*}, Karin Gaudenz², Robert Peuß², Jenny Y. Sung², Brian Martineau¹, Megan Peavey¹, Andrew C. Box², Julius A. Tabin¹, Suzanne McGaugh³, Richard Borowsky⁴, Clifford J. Tabin¹ & Nicolas Rohrer^{2,5}

Periodic food shortages are a major challenge faced by organisms in natural habitats. Cave-dwelling animals must withstand long periods of nutrient deprivation, as—in the absence of photosynthesis—caves depend on external energy sources such as seasonal floods¹. Here we show that cave-adapted populations of the Mexican tetra, *Astyanax mexicanus*, have dysregulated blood glucose homeostasis and are insulin-resistant compared to river-adapted populations. We found that multiple cave populations carry a mutation in the insulin receptor that leads to decreased insulin binding *in vitro* and contributes to hyperglycaemia. Hybrid fish from surface-cave crosses carrying this mutation weigh more than non-carriers, and zebrafish genetically engineered to carry the mutation have increased body weight and insulin resistance. Higher body weight may be advantageous in caves as a strategy to cope with an infrequent food supply. In humans, the identical mutation in the insulin receptor leads to a severe form of insulin resistance and reduced lifespan. However, cavefish have a similar lifespan to surface fish and do not accumulate the advanced glycation end-products in the blood that are typically associated with the progression of diabetes-associated pathologies. Our findings suggest that diminished insulin signalling is beneficial in a nutrient-limited environment and that cavefish may have acquired compensatory mechanisms that enable them to circumvent the typical negative effects associated with failure to regulate blood glucose levels.

The fish species *A. mexicanus* consists of infertile river-dwelling and cave-dwelling populations (Fig. 1a), here referred to as ‘surface fish’ and ‘cavefish’, respectively, that experience markedly different nutrient availability². Cavefish are resistant to starvation; when food-deprived, cavefish lose a smaller fraction of their body weight compared to surface fish³. Several factors have been identified that contribute to starvation resistance, including reduced metabolic circadian rhythm⁴, decreased metabolic rate⁵ and elevated body fat³. The genetic changes underlying these adaptations remain largely unknown.

In this study, we focused on three cavefish populations, which are named for the caves they inhabit (Tinaja, Pachón and Molino), that have evolved independently from two different stocks of surface fish that invaded caves millions of years ago⁶. Tinaja and Pachón cavefish originated from a more ancient surface population compared to Molino cavefish⁷.

A critical aspect of metabolic homeostasis is blood glucose regulation⁸. We compared blood glucose levels of Tinaja, Pachón and Molino cavefish with those of laboratory-raised surface fish, and found that the cave populations had significantly higher postprandial blood glucose levels (means of 64, 76 and 92 versus 47 mg dl⁻¹, respectively; Fig. 1b). We investigated the dynamics of glucose homeostasis during short- and long-term fasting (Fig. 1c). Cavefish had significantly higher blood glucose levels after 24 hours of fasting (mean of 80, 80 and 107 mg dl⁻¹ for Tinaja, Pachón and Molino populations, respectively, versus 59 mg dl⁻¹ for surface fish, $P < 0.05$, one-way ANOVA). After

21 days, we observed a marked decrease in blood glucose levels in Tinaja and Pachón cavefish; at 1 and 21 days, respectively, Tinaja cavefish showed mean blood glucose levels of 80 versus 48 mg dl⁻¹ ($P < 0.0001$) and Pachón cavefish showed mean levels of 80 versus 44 mg dl⁻¹ ($P < 0.0001$). Surface fish experienced a minor decrease from 59 to 40 mg dl⁻¹ (at 1 and 21 days, respectively, $P = 0.02$), and there was no significant change in Molino cavefish (107 versus 100 mg dl⁻¹ at 1 and 21 days, respectively, $P = 0.89$). Molino cavefish maintain elevated blood glucose levels, highlighting what appears to be a fundamental difference between their metabolic adaptation and the adaptations of the other cavefish populations that we investigated. However, our results suggest that dysregulated glucose homeostasis is a common feature of cavefish populations.

To further test this hypothesis, we compared acute control of glucose homeostasis using a glucose tolerance test (Fig. 1d). We injected glucose into the intraperitoneal cavity of surface fish and cavefish, transiently raising blood glucose levels to over 400 mg dl⁻¹ in most fish (Fig. 1d). Eight hours after injection, blood glucose levels of the surface fish were the same as those of PBS-injected controls (mean 126 versus 120 mg dl⁻¹, respectively; Fig. 1d) whereas blood glucose levels remained highly elevated in all cavefish populations (mean 374, 432 and 411 mg dl⁻¹ in Tinaja, Pachón and Molino populations, respectively, $P < 0.0005$, one-way ANOVA compared to PBS-injected controls; Fig. 1d). In Tinaja and Pachón cavefish, blood glucose levels remained elevated at 13.5 h after injection, but were not significantly different from those of PBS-injected controls (Tinaja cavefish, mean 157 versus 79 mg dl⁻¹ in glucose-injected and PBS-injected fish, respectively, $P = 0.47$; in Pachón cavefish, mean 159 versus 105 mg dl⁻¹ in glucose-injected and PBS-injected fish, respectively, $P = 0.65$). In Molino cavefish, blood glucose remained significantly elevated at 13.5 h (315 mg dl⁻¹ versus 144 mg dl⁻¹ in PBS-injected controls, $P = 0.000001$) and returned to the same levels as PBS-injected controls after 24 h (132 versus 133 mg dl⁻¹ in glucose-injected versus PBS-injected controls, respectively). Our results suggest that cavefish have impaired glucose clearance.

Glucose homeostasis requires the balanced release of insulin and glucagon that instructs tissues to absorb glucose from the blood or produce glucose from stored glycogen⁹. We found that at ten days post fertilization the first pancreatic islet is visible and does not differ in the number of cells producing glucagon (54 versus 50 in surface fish and cavefish, respectively, $P = 0.678$) or insulin (54 versus 52 in surface fish and cavefish, respectively, $P = 0.275$), which suggests that the pancreas develops similarly in surface fish and cavefish ($n = 5$ per population; Extended Data Fig. 1). In 1–2-year-old adults, we did not detect a difference in circulating glucagon levels after a 24 h fast (Extended Data Fig. 2). Circulating insulin levels tended to be higher in Tinaja cavefish, but the results were not significant ($n = 24$ per population; Extended Data Fig. 3). Nonetheless, we found evidence of diminished insulin response in cavefish. We injected the fish with arginine, which

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. ³College of Biomedical Sciences, University of Minnesota, St. Paul, Minnesota 55108, USA. ⁴Department of Biology, New York University, New York, New York 10003, USA. ⁵Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, Kansas 66160, USA.

*These authors contributed equally to this work.

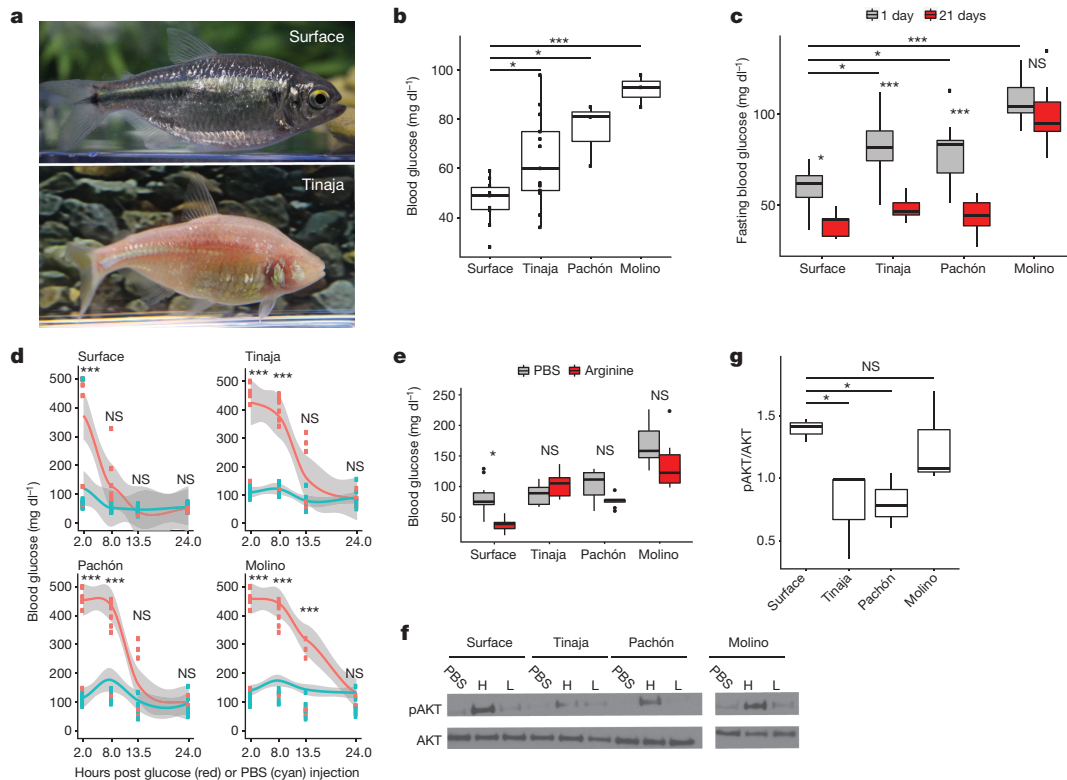


Figure 1 | Altered glucose homeostasis in cave-adapted *A. mexicanus* populations. **a**, Surface fish and Tinaja cavefish of *A. mexicanus*. **b**, Blood glucose (1 h postprandial) in surface fish compared to cavefish ($n = 10$, 13, 3 and 3, respectively, for surface fish, Tinaja, Pachón and Molino cavefish). **c**, Fasting blood glucose at day 1 versus day 21 ($n = 20$ per population and condition). **d**, Glucose tolerance test. Blood glucose after intraperitoneal injection of glucose (red) or PBS (blue). Data points represent values for individual fish and grey shade indicates 95% confidence interval for polynomial regression. **e**, Blood glucose 5 h after intraperitoneal injection of arginine ($n = 10$ per population and

condition). **f**, Western blot: cell lysates probed with pAKT (ser473) and AKT antibodies. Lysates produced from skeletal muscle treated *ex vivo* with PBS, a high (H, 9.5–11.5 $\mu\text{g ml}^{-1}$) or a low (L, 0.95–1.15 $\mu\text{g ml}^{-1}$) level of insulin. **g**, Quantification of bands by densitometry of highest concentration treatment ($n = 3$ per population). For box plots, median, 25th, 50th and 75th percentiles are represented by horizontal bars, and vertical bars represent 1.5 \times interquartile ranges. Significance calculated using one-way ANOVA with Tukey's honest significant difference (HSD) post hoc test. NS, $P > 0.05$; * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0005$. For gel source data, see Supplementary Fig. 1.

stimulates the simultaneous release of glucagon and insulin¹⁰, and observed that while surface fish experienced a significant decrease in blood glucose level (mean 80 versus 38 mg dl⁻¹, PBS-injected versus arginine-injected fish, $P = 0.006$), cavefish blood glucose levels did not change (Fig. 1e). In addition, we found that injection of recombinant human insulin caused a significant drop in blood glucose after 60 min in surface fish, but not in Tinaja cavefish (Extended Data Fig. 4). Our combined observations that glucagon and insulin levels do not differ between surface and cavefish, and that cavefish do not decrease blood glucose levels in response to arginine or insulin, suggest that cavefish may be insulin resistant.

Insulin-stimulated glucose uptake proceeds through phosphorylation of AKT at serine 473 (pAKT)¹¹. We compared the ratio of pAKT to AKT in freshly dissected skeletal muscle treated with recombinant insulin (Fig. 1f, g). Consistent with the apparent dysregulation of glucose homeostasis, we observed lower pAKT levels in Tinaja cavefish than in surface fish (mean 0.775 versus 1.39 pAKT:AKT ratio, respectively, $P = 0.017$; Fig. 1f, g), which suggests that Tinaja cavefish are indeed insulin resistant relative to surface fish. Pachón cavefish muscle also had lower pAKT levels in response to insulin (mean 0.806 pAKT:AKT ratio, $P = 0.027$; Fig. 1f, g); however, despite having elevated blood glucose (Fig. 1b), Molino cavefish pAKT levels were equivalent to surface fish (mean 1.26 pAKT:AKT ratio, $P = 0.99$; Fig. 1f, g). Our results suggest that Tinaja and Pachón fish have evolved altered blood glucose regulation and insulin resistance in parallel, whereas Molino fish may have evolved their altered glucose metabolism through a different mechanism.

To investigate the genetic mechanism underlying insulin resistance in cavefish, we examined the sequences of all known genes in the

insulin pathway using the available genome sequence¹² (Supplementary Information 1). Notably, we found a coding difference in the insulin receptor gene (*insra*) between surface fish and cavefish that affects a conserved proline in the extracellular cysteine-rich domain (P211L; Fig. 2a–c). The presence of the mutation correlates with insulin resistance, as Tinaja and Pachón populations carry the mutation whereas Molino cavefish have the wild-type allele (Fig. 2b, c). Notably, the same genetic alteration is implicated in at least two known cases of Rabson–Mendenhall syndrome^{13,14}, a form of severe insulin resistance in humans (Fig. 2c). The biochemical effect of the mutation has not previously been explored, but the position in the cysteine-rich domain suggests a role in insulin binding¹⁵. To test this, we generated transgenic HEK293T (Flp-In-293) cell lines that stably express the full-length surface fish or Tinaja cavefish *insra* and incubated the cells with different concentrations of fluorescein isothiocyanate (FITC)-labelled human insulin. We measured fluorescence as a readout for binding efficiency using an image-based cytometry approach (Imagestream X Mark II) and found that cells that expressed the cavefish receptor displayed significantly lower binding at all but the lowest concentrations of insulin (Fig. 2d). Our results suggest that the P211L mutation of *insra* affects insulin signalling by altering insulin binding efficiency.

We next tested for the presence and frequency of the P211L mutation in wild-caught fish. We genotyped 71 surface fish from different localities, and 51 cavefish from 6 different caves (Supplementary Information 2; Fig. 3a). Consistent with our observations above, the mutation was absent in Molino fish ($n = 8$) but present in all other tested cave populations (Tinaja, Yerbániz, Pachón, Japonés and Arroyo, combined $n = 36$). Notably, the cave populations carrying this mutation are all

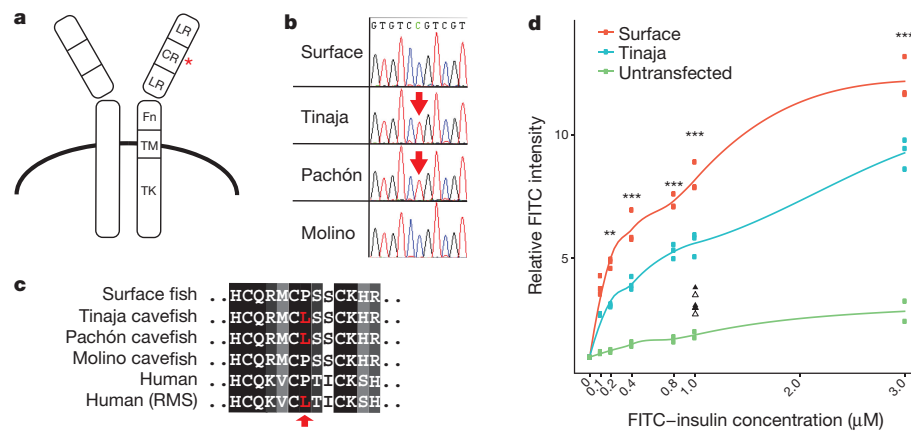


Figure 2 | Coding mutation in the cavefish insulin receptor leads to decreased insulin binding. **a**, Schematic of the insulin receptor (adapted from ref. 24). Red asterisk depicts position of the P211L mutation. LR, leucine-rich repeats; CR, cysteine-rich domain; Fn, fibronectin type III domain; TM, transmembrane domain; TK, tyrosine-kinase domain. **b**, Sequence chromatogram of the mutation in *Astyanax*. **c**, Amino acid alignment of the insulin receptor P211L mutation with patients with Rabson-Mendenhall syndrome ('Human (RMS)'). **d**, Relative FITC intensity of cells stably transfected with Flag-tagged surface-fish or

Tinaja-cavefish insulin receptor and incubated with FITC-labelled insulin. Each point represents mean FITC intensity of >2,500 live cells normalized to the mean intensity of untreated cells. Lines represent results from local polynomial regression fitting. Triangles (surface fish, filled; Tinaja cavefish, unfilled) represent data from competitive binding assay in which cells were incubated with 10 μ M unlabelled insulin. Significance calculated using one-way ANOVA (between surface fish and Tinaja cavefish) with Tukey's HSD post hoc test, * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0005$.

derived from the same ancestral stock of surface fish¹⁶. Although the mutation was present in heterozygote conditions in some of the caves, we did not find any cavefish homozygous for the surface allele (Fig. 3a). Our findings suggest that there is active selection for the mutation in the caves, and a partially dominant effect of the cave allele. We did not observe the cave allele in any of the surface fish, which suggests that the mutation either appeared *de novo* in the cave populations, represents a rare variant not detected by our sampling frequency or is absent in the current surface population but was present in the ancestral surface fish stocks¹⁶ (Supplementary Information 3).

To investigate whether the P211L mutation contributes to altered glucose regulation, we measured fasting blood glucose levels in 192 F₂ fish from a surface fish–Tinaja cavefish cross fed *ad libitum* their entire lives (Extended Data Fig. 5). We found that elevated blood glucose—defined as levels that exceed the surface fish mean (59 mg dl⁻¹) or maximum (75 mg dl⁻¹)—is a non-Mendelian trait in our cross: only 4.7%, rather than the expected 25%, of fish inherited elevated blood glucose, which indicates that the trait is not monogenic. Although F₂ fish with the P211L mutation did not differ significantly in blood glucose levels from those without the mutation, we found that only F₂ fish carrying the P211L mutation have elevated blood glucose levels exceeding the surface fish mean ($n = 9$) or maximum ($n = 3$), which suggests a necessary but not sufficient role for the mutation in altered blood glucose regulation.

We found that cavefish weigh more than surface fish on a nutrient-limited diet (mean 2.08 versus 1.52 g, respectively, $P = 0.02$; Fig. 3b). To investigate whether the *insra* mutation influences weight, we genotyped and weighed 124 surface–Tinaja male F₂ fish at approximately 1.5 years of age that were fed *ad libitum*. We focused on males, as egg mass varies between individual females and can account for as much as 41% of female body weight (Extended Data Fig. 6), representing a confounding variable if included in our analysis. We found that males carrying one or two copies of the cave P211L *insra* allele weigh on average 27% more than hybrids that carry only the surface allele (mean 1.63 versus 1.28 g, respectively, $P = 0.006$; Fig. 3c). Cavefish have an increased appetite that is associated with a mutation in the melanocortin 4 receptor³. Although this mutation segregates independently from the *insra* mutation in the F₂ hybrid population, it is possible that another locus in *cis* to the P211L mutation influences appetite. To eliminate the effect of appetite on our analysis of the *insra* allele, we housed F₂ fish of approximately the same starting weight (<2 g) individually and ensured they ate a diet of 6 mg of food per day for 4 months. We

found that homozygous P211L fish gained significantly more weight (mean 0.37 g, $n = 20$) than fish that did not carry the cave allele (0.19 g, $n = 60$, $P = 0.02$) consistent with the idea that variation at the *insra* locus influences weight gain independent of appetite regulation.

To determine whether the insulin resistance and weight gain associated with the P211L mutation in cavefish are indeed due to alteration of the *insra* gene, we used CRISPR gene editing to introduce the mutation into zebrafish (*Danio rerio*) via homology directed repair¹⁷ (Extended Data Fig. 7). We found that zebrafish homozygous for the P211L mutation have a lower ratio of pAKT to AKT in their skeletal muscle compared to heterozygous siblings, in both untreated (0.05 versus 0.13 in P211L-homozygous and heterozygous fish, respectively, $P = 0.016$) and insulin-stimulated conditions (0.13 versus 0.32 in P211L-homozygous and heterozygous fish, respectively, $P = 0.067$, $n = 3$ per genotype and condition) (Fig. 3f). In addition, we found that zebrafish homozygous for the cave allele are longer (mean 20.0 mm, versus 18.4 mm in heterozygous fish, $P = 0.0046$) and weigh more (mean 124.6 mg, versus 99.7 mg in heterozygous fish, $P = 0.022$) than their siblings when raised under the same conditions (Fig. 3e, g, h). Our findings show that the P211L mutation contributes to both the increased weight and insulin resistance observed in Tinaja and Pachón cavefish. Increased body weight in fish homozygous for the insulin-receptor mutation is unexpected, considering that in mammals, full loss-of-function mutations in the insulin receptor are associated with retarded growth and lower levels of body fat¹⁸. Our results suggest that diminished insulin signalling has an opposite effect in fish, but the mechanisms leading to the difference remain unclear.

Cavefish are insulin resistant and hyperglycaemic; in humans these phenotypes precede and define type 2 diabetes, respectively. In addition, cavefish have a fatty liver³, which is also frequently associated with type 2 diabetes. Transgenic zebrafish carrying the *insra* mutation show a reduction in scale size (Extended Data Fig. 8), as previously described in other hyperglycaemic zebrafish models¹⁹ and a similar phenotype has recently been described in cavefish²⁰. In principle, these data could suggest an evolutionary tradeoff in which physiological health is sacrificed to reap the benefits of starvation resistance. However, whereas surface fish begin to exhibit signs of ageing such as sunken skin, tattered fins and a hunched back²¹ (Fig. 4a) by age 15, Tinaja and Pachón cavefish can live in excess of 14 years without these indications of senescence (Fig. 4b, c) and with no difference in fertility decline relative to surface fish. Cavefish may have evolved compensatory mechanisms enabling them to remain healthy despite potentially deleterious metabolic changes.

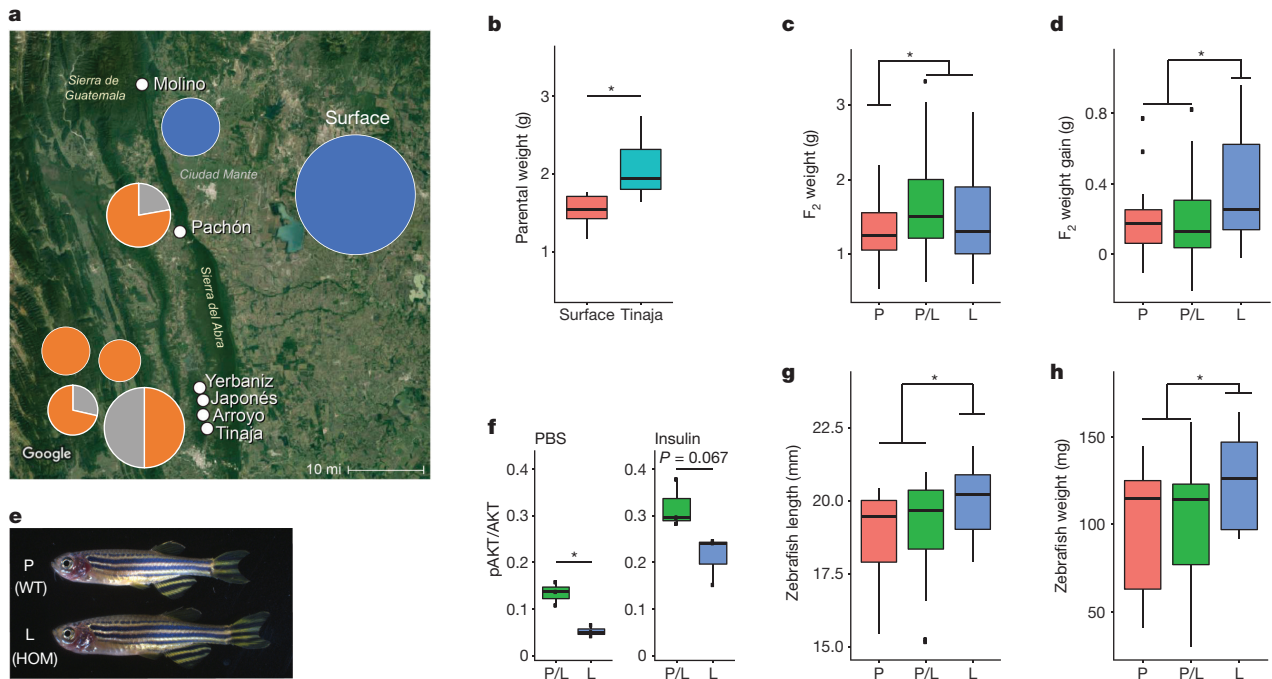


Figure 3 | The P211L mutation of *insra* is overrepresented in cave environments and is associated with higher body weight in surface-cave hybrids. **a**, Map of the region, overlain with genotyping results of wild-caught samples. Pie charts indicate percentage of fish homozygous for surface allele (blue), cave allele (orange) or heterozygous (grey). Size of pie chart roughly indicates the number of fish genotyped (Molino, $n = 8$; Surface, $n = 71$; Pachón, $n = 9$; Yerbaniz, $n = 8$; Japonés, $n = 5$; Arroyo, $n = 7$; Tinaja, $n = 14$; position of pie charts corresponds to location and vertical order of population name on the map). The P211L allele is absent in all wild-caught surface fish and Molino cavefish (descended from a more-recent surface-fish lineage). The P211L mutation is present in all sampled cavefish populations descended from the more-ancient surface-fish lineage. Tinaja, Yerbaniz, Japonés and Arroyo are geographically close and believed to represent a single invasion event; Pachón represents an independent invasion²⁵. Map source: Imagery ©2017 Landsat/Copernicus,

Map data ©2017 Google, INEGI. **b**, Weight of Tinaja males ($n = 6$) and surface males ($n = 5$) on a nutrient-limited diet. **c**, Weight of 18-month-old F₂ male Tinaja-surface hybrids genotyped for the P211L mutation. P-homozygous (P) surface fish, $n = 22$; L-homozygous (L) cavefish, $n = 27$; heterozygotes (P/L), $n = 53$. **d**, Change in weight of F₂ Tinaja-surface hybrid males on fixed diet. $n = 21$ (P), 39 (P/L) and 20 (L). **e**, Images of wild-type (WT) and homozygous P211L mutant (HOM) zebrafish siblings. **f**, Ratio of pAKT:AKT in adult zebrafish skeletal muscle treated *ex vivo* with PBS or insulin ($n = 3$ per genotype and condition). **g**, **h**, Length and weight of wild-type zebrafish ($n = 13$ (P)) and heterozygous ($n = 22$ (P/L)) and homozygous ($n = 11$ (L)) P211L mutant zebrafish. In box plots the median, 25th, 50th, and 75th percentiles are represented by horizontal bars and vertical bars represent $1.5 \times$ interquartile ranges. Significance calculated using two-tailed students *t*-test, * $P < 0.05$.

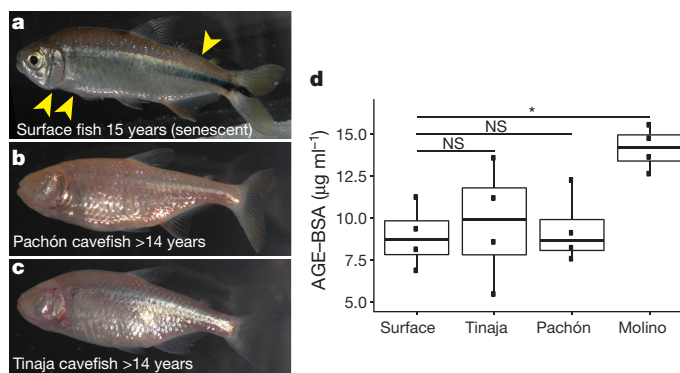


Figure 4 | Despite elevated blood glucose levels and insulin resistance, Tinaja and Pachón cavefish do not show signs of senescence and do not accumulate advanced glycation end-products in the blood. **a–c**, Surface (a), Pachón (b) and Tinaja (c) fish kept in the laboratory for the indicated duration and fed *ad libitum*. Cavefish were wild-caught; ages represent minimum age. Surface fish (a) shows signs of ageing, such as loose skin and bent tails (yellow arrows), that are absent in cavefish at comparable ages (b, c). **d**, Quantification of advanced glycation end-products in serum (AGE-BSA) from approximately two-year-old fish after a three-day fast ($n = 4$ for each population). * $P < 0.05$, one-way ANOVA with Tukey's HSD post hoc test. For box plots, median, 25th, 50th and 75th percentiles are represented by horizontal bars and vertical bars represent $1.5 \times$ interquartile ranges.

A major cause of morbidity in diabetic patients is tissue damage caused by excessive non-enzymatic glycation of proteins in the blood, generating advanced glycation end-products (AGEs)²². AGEs are closely associated with diabetes-induced vascular damage, cardiovascular disease and ageing²³. We compared the level of AGEs in the serum of two-year-old fish that had been fed *ad libitum* their entire lives (Fig. 4b). We did not detect a difference in the levels of AGEs between Tinaja and Pachón cavefish relative to surface fish (mean of 9.7, 9.3 and 8.9 $\mu\text{g ml}^{-1}$, respectively, $P = 0.99, 0.95$), despite the elevated blood glucose levels of the cavefish (Fig. 1b). These two populations may have mechanisms for reducing protein glycation, rendering them impervious to the damaging effects of elevated blood glucose. Notably, the Molino cavefish do have elevated levels of AGEs (mean 14.1 $\mu\text{g ml}^{-1}$, versus 8.9 $\mu\text{g ml}^{-1}$ in surface fish, $P = 0.03$). It remains to be determined whether the health and longevity of the Molino population is influenced by accumulation of AGEs, but our results suggest they may have evolved altered blood glucose homeostasis through a different mechanism than did Tinaja and Pachón cavefish.

Our findings establish cavefish as a model with which to investigate resistance to pathologies of diabetes-like dysregulation of glucose homeostasis. Moreover, our results highlight the extreme physiological measures that can evolve in critical metabolic pathways to accommodate exceptional environmental challenges.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 August 2017; accepted 15 February 2018.

Published online 21 March 2018.

1. Culver, D. C. & Pipan, T. *The Biology of Caves and Other Subterranean Habitats* (Oxford Univ. Press, 2009)
2. Horst Wilkens, U. S. *Evolution in the Dark, Darwin's Loss Without Selection* (Springer, 2017)
3. Aspiras, A. C., Rohnner, N., Martineau, B., Borowsky, R. L. & Tabin, C. J. Melanocortin 4 receptor mutations contribute to the adaptation of cavefish to nutrient-poor conditions. *Proc. Natl Acad. Sci. USA* **112**, 9668–9673 (2015).
4. Moran, D., Softley, R. & Warrant, E. J. Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism. *PLoS ONE* **9**, e107877 (2014).
5. Hüppop, K. Oxygen consumption of *Astyanax fasciatus* (Characidae, Pisces): a comparison of epigeal and hypogeal populations. *Environ. Biol. Fishes* **17**, 299–308 (1986).
6. Gross, J. B. The complex origin of *Astyanax* cavefish. *BMC Evol. Biol.* **12**, 105 (2012).
7. Bradic, M., Teotónio, H. & Borowsky, R. L. The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Mol. Biol. Evol.* **30**, 2383–2400 (2013).
8. Saltiel, A. R. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799–806 (2001).
9. Rines, A. K., Sharabi, K., Tavares, C. D. & Puigserver, P. Targeting hepatic glucose metabolism in the treatment of type 2 diabetes. *Nat. Rev. Drug Discov.* **15**, 786–804 (2016).
10. Navarro, I. et al. Insights into insulin and glucagon responses in fish. *Fish Physiol. Biochem.* **27**, 205–216 (2002).
11. Lizcano, J. M. & Alessi, D. R. The insulin signalling pathway. *Curr. Biol.* **12**, R236–R238 (2002).
12. McGaugh, S. E. et al. The cavefish genome reveals candidate genes for eye loss. *Nat. Commun.* **5**, 5307 (2014).
13. Atray, A. et al. Rabson Mendenhall Syndrome; a case report. *J. Diabetol.* **2**, 2 (2013).
14. Carrera, P. et al. Substitution of Leu for Pro-193 in the insulin receptor in a patient with a genetic form of severe insulin resistance. *Hum. Mol. Genet.* **2**, 1437–1441 (1993).
15. Taylor, S. I. et al. Mutations in insulin-receptor gene in insulin-resistant patients. *Diabetes Care* **13**, 257–279 (1990).
16. Bradic, M., Beerli, P., García-de León, F. J., Esquivel-Bobadilla, S. & Borowsky, R. L. Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*). *BMC Evol. Biol.* **12**, 9 (2012).
17. Albadri, S., Del Bene, F. & Revenu, C. Genome editing using CRISPR/Cas9-based knock-in approaches in zebrafish. *Methods* **121–122**, 77–85 (2017).
18. Savage, D. B. & Semple, R. K. Recent insights into fatty liver, metabolic dyslipidaemia and their links to insulin resistance. *Curr. Opin. Lipidol.* **21**, 329–336 (2010).
19. Suzuki, N., Kitamura, K. I. & Hattori, A. Fish scale is a suitable model for analyzing determinants of skeletal fragility in type 2 diabetes. *Endocrine* **54**, 575–577 (2016).
20. Simon, V. et al. Comparing growth in surface and cave morphs of the species *Astyanax mexicanus*: insights from scales. *Evodevo* **8**, 23 (2017).
21. Hayes, A. J. et al. Spinal deformity in aged zebrafish is accompanied by degenerative changes to their vertebrae that resemble osteoarthritis. *PLoS ONE* **8**, e75787 (2013).
22. Yan, S. F., Ramasamy, R. & Schmidt, A. M. Mechanisms of disease: advanced glycation end-products and their receptor in inflammation and diabetes complications. *Nat. Clin. Pract. Endocrinol. Metab.* **4**, 285–293 (2008).
23. Prasad, A., Bekker, P. & Tsimikas, S. Advanced glycation end products and diabetic cardiovascular disease. *Cardiol. Rev.* **20**, 177–183 (2012).
24. De Meyts, P. & Whittaker, J. Structural biology of insulin and IGF1 receptors: implications for drug design. *Nat. Rev. Drug Discov.* **1**, 769–783 (2002).
25. Borowsky, R. Restoring sight in blind cavefish. *Curr. Biol.* **18**, R23–R24 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Chinchore and C. Sengel for technical advice; X. Gao for bioinformatics support; Z. Zakibe for photographs of the fish; the Aquatics facility at Stowers for fish maintenance and support; the cell culture core at Stowers for cell line maintenance and advice; the molecular biology core at Stowers for design, execution and validation of the CRISPR constructs; the proteomics core; M. Levy for advice and computational modelling of the insulin receptor; A. Herman for help with the genome scan; the Microscopy Resources on the North Quad (MicRoN) core at Harvard Medical School; M. Miller for illustration; and S. Williams, F. Damen, S. Xiong, E. Kingsley and K. Fox for feedback on the manuscript text. This work was supported by a grant from the NIH to C.J.T. (HD089934) and institutional funding to N.R. M.R.R. was supported by a National Research Service Award (DK108495) and R.P. was supported by a grant from the Deutsche Forschungsgemeinschaft (PE 2807/1-1).

Author Contributions M.R.R., A.C.A., C.J.T. and N.R. conceived the project and designed research with additional contributions from K.G., R.P. and A.C.B. M.R.R., A.C.A., K.G., R.P., J.Y.S., B.M., M.P., A.C.B., J.A.T., S.M., R.B. and N.R. performed the research. M.R.R., A.C.A., C.J.T. and N.R. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to N.R. (nro@stowers.org) or C.J.T. (tabin@genetics.med.harvard.edu).

Reviewer Information *Nature* thanks K. Kavanagh, S. O'Rahilly and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. Investigators were blinded to fish genotypes during experiments when possible.

Fish husbandry and diet. Fish husbandry was performed as previously described²⁶. Unless stated otherwise, fish were fed *ad libitum* with a combination of New Life Spectrum TheraA + small fish formula and *Artemia* and housed at densities of less than or equal to two adult fish per litre of water. F₂ hybrids were housed individually in 1.5 l tanks and fed three pellets (~6 mg) of New Life Spectrum TheraA + small fish formula once per day for over 4 months. For the starvation experiment, fish were moved to individual containers and water was changed daily. **Blood glucose, glucose tolerance and arginine tolerance.** Blood was collected from the caudal tail vein using a U-100 insulin needle and glucose was measured using FreeStyle Lite blood glucose meter and test strips.

Glucose (2.5 mg/gram fish), arginine (6.6 µM/gram fish), or PBS was injected into the intraperitoneal cavity using a U-100 insulin needle.

pAKT quantification. We quantified pAKT level in fillets of skeletal muscle taken directly after fish decapitation. For *A. mexicanus*, skeletal muscle was cut into three equal strips per fish. Strips were incubated in PBS, 0.1 × or 1 × concentration of recombinant human insulin for 25 min (Sigma product I9278; 1 × concentration = 9.5–11.5 µg/ml insulin). The tissues were rinsed in PBS and then homogenized and lysed in RIPA buffer (Sigma) with protease and phosphatase inhibitor (Pierce Protease and Phosphatase Inhibitor Mini Tablets, EDTA Free) for 30 min. Protein concentration was measured via BCA (Pierce). Lysate protein concentrations were then equalized and run on 4–12% Bis-Tris protein gel and transferred on a nitrocellulose membrane. Blots were probed for AKT (Cell Signaling). Following stripping, blots were probed for pAKT (ser473) (Cell Signaling). ImageJ was used for densitometry measurements. For *D. rerio* two fillets of skeletal muscle were removed from both sides of fish directly after decapitation. Fillets were rinsed in PBS and then incubated in PBS or 10 µg/ml human recombinant insulin (Sigma, I0908) for approximately 40 min. The skin was then removed from the skeletal muscle and the muscle was finely minced using a scalpel. We quantified the ratio of pAKT:AKT using the Akt(pS473) + total Akt ELISA Kit (Abcam ab126433) according to the manufacturers' protocol. We used 200 µl lysis buffer per sample and loaded 85 µl of lysate per well.

insra P211L genotyping in *A. mexicanus*. Genomic DNA from tail fin clips was diluted fivefold and used as target DNA to amplify the *insra* locus using the following oligonucleotide primers: *insra*_f: GCACCCTTACACCCTTACATGA; *insra*_r: TACCGCTCAGCACTAATTGGA; product size: 700 bp. PCR reactions were carried out in 12.5-µl volume containing 1 × LA PCR Buffer II (Clontech), 2.5 mM MgCl₂, 0.4 mM dNTP mix, 0.4 µM of each forward and reverse primer and 0.05 units of TaKaRa LA Taq DNA Polymerase (Clontech). The PCR cycling conditions were as follows: initial denaturation at 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, annealing temperature 52 °C for 30 s and 72 °C for 1 min. A final 5-min elongation step was performed at 72 °C. The PCR products were diluted tenfold and sequenced directly on a 3730XL DNA Analyzer (Applied Biosystems) using the sequencing primer: GGTGGAGTTGATGGTGGTATAG.

Selection scans at the *insra* locus. We examined the *insra* locus with data that are a part of an ongoing genome-wide selection and demography companion study (A. Herman *et al.*, manuscript in preparation). Methods are explained in greater detail in this demographic companion study; in brief, we used Illumina HiSeq 2000 to sequence 100-bp reads from 6–10 individuals from each population of Tinaja cave, Molino cave, Pachón cave, Rascon surface and Río Choy surface populations, (total *n* = 43) and two individuals from the sister taxa *Astyanax aeneus*. Individuals were sequenced with v3 chemistry. Reads were cleaned with Trimmomatic v.0.30²⁷ and cut-adapt v.1.2.1²⁸ and aligned to the reference Pachón cavefish genome using bwa-mem algorithm in bwa-0.7.1²⁹ resulting in an aligned coverage depth of ~7–12 ×. Variants were called using the Genome Analysis Toolkit v.3.3.0 (GATK)³⁰ and Picard v.1.83 (<http://broadinstitute.github.io/picard/>). Outlier scan metrics (π , DXY, FST and Tajima's *D*) were conducted using VCFtools v.0.1.13³¹ and custom scripts. HSCAN (<https://messerlab.org/resources/>) and hapFLK³² were also used to examine *insra* for outliers. Metrics were dense-ranked across the genome and the ranking position of *insra* was used to determine whether it was exceptionally divergent between cave and surface populations relative to the rest of the genome.

Glucagon and insulin quantification. The number of cells producing insulin or glucagon was determined using the following protocol: 10–11 days post fertilization fish were euthanized with an overdose of tricaine and fixed in 4% paraformaldehyde overnight at 4 °C. Fish were washed in PBST, transferred to water for 1 min, acetone for 10 min at –20 °C, water for 1 min, and then washed in PBST. Blocking in 5% donkey serum, 1% DMSO and 0.2% BSA was done for 1 h at room temperature. Fish were incubated with primary antibodies (1:200 sheep anti-glucagon (Abcam), 1:200 guinea pig anti-insulin (DAKO)) and then secondary antibodies (1:400 donkey anti-sheep 488, 1:400 goat anti-guinea pig 647) overnight

at room temperature in glass vials, then washed with PBST, stained with DAPI and imaged. Images were collected at 63 × using a 1.0 µm z-stack on a Zeiss 780 confocal microscope. Nuclei surrounded by insulin or only glucagon were counted manually using Fiji cell counter.

To quantify circulating glucagon level, we collected serum from the caudal tail vein of fish that were approximately 2-years-old and had fasted for 24 h (*n* = 12 for each population). The serum was used for a glucagon radioimmunoassay according to the manufacturer's protocol (MGL-32K; Millipore). To quantify circulating insulin levels, we collected plasma from the caudal tail vein of 2-year-old and 1-year-old fish and blotted the serum onto a nitrocellulose membrane using a Bio-dot sf device (Bio-Rad). Ponceau protein staining was used to verify equal protein loading, after which insulin was probed using anti-insulin (DAKO). Quantification of insulin levels was done using Fiji.

Insulin-binding experiment. The full-length *A. mexicanus insra* protein-coding sequence was amplified from surface-fish and Tinaja-cavefish cDNA and cloned into a modified pcDNA3.1/Hygro vector providing an N-terminal Flag epitope tag³³. To generate stable cell lines, the Flag-tagged *insra* cassettes were cloned into pcDNA5/FRT vector (Invitrogen, catalogue number V601020) allowing for FLP recombinase-mediated integration into the FLP-In-293 cell line according to the manufacturer's procedures. The FLP-In-293 cell line was originally purchased from Invitrogen/Thermo Fisher Scientific (catalogue number R75007), then cell-banked internally at the Stowers Institute. Cell line authentication was performed by Promega/ATCC, using short tandem repeat (STR) profiling. Cell lines were tested for mycoplasma contamination using the Universal Mycoplasma Detection Kit, ATCC 30-1012. One positive clone from each surface-fish and Tinaja-cavefish cell line was selected and used for the insulin-binding assays using a protocol derived from a previously published protocol³⁴. In brief, 100-mm plates were seeded at 30% confluency and cultured in DMEM, 10% FBS + 1 × Glx medium for 48 h. The plates were then changed to insulin-free FreeStyle 293 Expression Medium (catalogue number 12338018) and incubated for an additional 24 h. Plates at ~70–80% confluency were pre-chilled for 30 min at 4 °C and the medium was replaced with 5 ml of cold FreeStyle 293 Expression Medium containing 42 mM HEPES pH7.5 and human FITC-labelled insulin (Sigma, catalogue number I3661) at final concentrations of 0, 0.1, 0.2, 0.4, 0.8, 1 and 3 µM. For binding competition, 10 µM of unlabelled human insulin (Sigma, catalogue number I9278) and 1 µM of human FITC-labelled insulin were added. After one hour of incubation in the dark at 4 °C, the medium was removed and the plates were washed with 5 ml of cold 1 × PBS. Cells were dissociated in 2 ml of 0.5 mM EDTA in PBS at 37 °C for 7 min, transferred into Eppendorf tubes, pelleted for 5 min at 200g at 4 °C and resuspended in 1 ml of cold 1 × PBS. To stain dead cells, 1 µl of Fixable Viability Dye (FVD eFluor450, Invitrogen, catalogue number 65-0863) was added to each 1 ml of cell suspension and incubated on ice in the dark for 20 min. Subsequently, the cells were washed once with cold 1 × PBS and fixed in 1 ml of 4% formaldehyde. After two more washes with cold 1 × PBS, the cells were resuspended in 150 µl of cold 1 × PBS, filtered through a 70-µm cell strainer (Falcon 070-67-S) and transferred into a round-bottom 96-well plate. Binding data were acquired on an ImagestreamX MarkII (EMD Millipore) at 40 ×. FITC was excited with 150 mW 488 nm. Fixable live/dead was excited with 12 mW 405 nm. Single colour controls were used for colour compensation. Bright field was collected on different channels. Analysis was performed in IDEAS v.6.2 and fluorescence intensity was reported as integrated intensity within an adaptive erode mask for bright field.

Genome editing in zebrafish using the CRISPR–Cas9 technology. We introduced the cavefish P211L coding alteration into *D. rerio* by causing a precise C632T base pair exchange on the third exon of the *insra* gene (Extended Data Fig. 7). We designed the guide RNA target site using the web tool of the MIT Zhang laboratory (<http://crispr.mit.edu>). We then validated the target region and checked for naturally occurring SNPs by PCR and sequencing of genomic DNA using the following oligonucleotide forward, reverse and sequencing primers, respectively: (1) TGAGGTGTGTCGAGTGTCT; (2) TGTGCGTGTTCGTTGAGTGT, product size of 598 bp; and (3) CAGCCCTTGAAGGTGTAGAA. The PCR reaction components and procedures followed that described in '*insra* P211L genotyping in *A. mexicanus*'.

We used Cas9 protein from PNA Bio and 2-part Alt-R guide RNAs from IDT. Single-stranded oligodeoxynucleotides (ssODNs) were ordered as ultramers from IDT for generating the SNP mutations. The ssODNs included 100 bp of homology arms as well as silent SNP mutations in the guide RNA target site to prevent re-annealing of the guide RNA following homologous recombination. We also protected both ends of the ssODNs with three phosphorothioate bonds to inhibit exonuclease degradation in the cell. To form the guide RNA complex, we first annealed the specific CRISPR RNA (crRNA) with *trans*-activating crRNA (1 µM), followed by hybridization with the Cas9 protein (6.6 µM) to form a ribonucleoprotein complex. This mixture was incubated for 10 min. Next, we combined the ribonucleoprotein complex with the ssODNs (50 ng/µl) to form a total volume of

15 µl of injection mix, which was then injected into the cell at the 1-cell stage in at least 200 and up to 1,200 zebrafish embryos.

insra P211L genotyping in zebrafish. We designed a genotypic screening assay to confirm the expected location of the P211L SNP mutation at various points through development. First, after injections we divided the embryos into two groups; group 1 to screen at the embryo stage and group 2 to raise if the mutation was detected in group 1. Out of the 93 embryos screened, 34 fish exhibited a mixture of insertion and deletion mutations (35.5%), and at least 1 fish showed the exact SNP mutation (1.1%). Once these fish matured, we then identified adult mosaic individuals. Out of the 64 screened F₀ adult fish, 17 fish carried the intended P211L SNP mutation with the corresponding silent SNP changes (26.6%), but only 5 fish showed germline transmission (7.81%) when outcrossed to wild-type zebrafish. Next, we genotyped the F₁ embryos to calculate the germline transmission rate of each F₀ mosaic founder. The germline transmission rate ranged from 8.3% to 45.8% in the 5 F₀ fish that showed positive germline transmission. We then raised the progenies of these mosaic founders. Lastly, we genotyped these F₁ progenies upon maturity and maintained the line by outcrossing heterozygous mutants to wild-type fish in each subsequent generation. To test whether the P211L mutation contributes to zebrafish weight, we paired heterozygous fish with successful germline transmission and weighed their progeny shortly before maturity (55 days post fertilization) to avoid gonadal effects on weight.

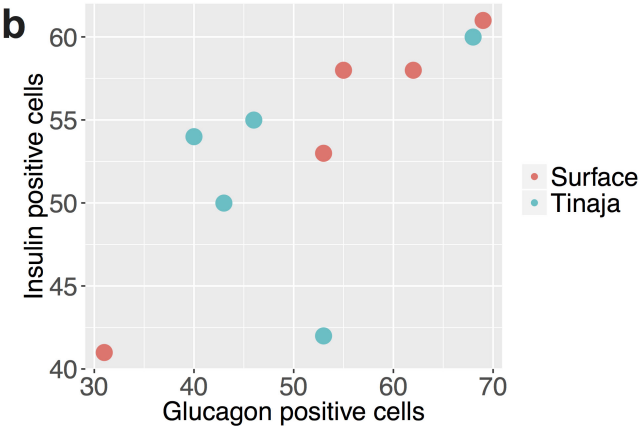
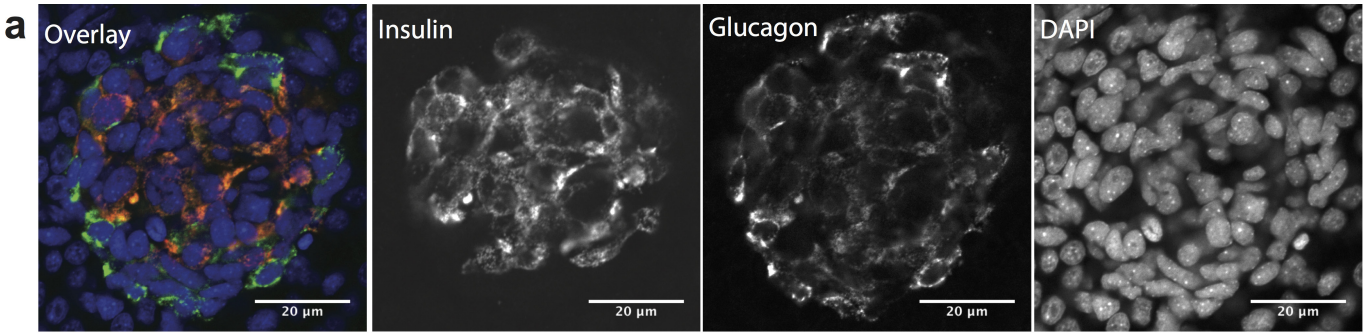
Quantification of advanced glycation end-products. We used Oxiselect Advanced Glycation End Product (AGE) Competitive ELISA Kit according to the manufacturer's protocol to measure AGE level in serum from two-year-old fish that were fasted for three days.

Statistics and figure preparation. Graphical data and statistics were produced using R³⁵ and ggplot2³⁶ package. We used Shapiro–Wilk to test for normality. We tested significance in normally distributed data using a *t*-test and non-parametric data using Mann–Whitney *U*-test. For non-parametric data with multiple ties we used a *t*-test. For comparison of more than two groups we used a one-way ANOVA with Tukey's HSD post hoc test.

Animal experiment statement. Research and animal care were approved by the Institutional Animal Care and Use Committees of the laboratories involved.

Data availability. Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at <http://www.stowers.org/research/publications/libpb-1205> and/or are available from the corresponding authors on reasonable request.

26. Elipot, Y., Legendre, L., P  re, S., Sohm, F. & R  taux, S. *Astyanax* transgenesis and husbandry: how cavefish enters the laboratory. *Zebrafish* **11**, 291–299 (2014).
27. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
28. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, <http://dx.doi.org/10.14806/ej.17.1.200> (2011).
29. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
30. Van der Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33 (2013).
31. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
32. Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M. & Servin, B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**, 929–941 (2013).
33. Tomomori-Sato, C. *et al.* A mammalian mediator subunit that shares properties with *Saccharomyces cerevisiae* mediator subunit Cse2. *J. Biol. Chem.* **279**, 5846–5851 (2004).
34. Murphy, R. F., Powers, S., Verderame, M., Cantor, C. R. & Pollack, R. Flow cytometric analysis of insulin binding and internalization by Swiss 3T3 cells. *Cytometry* **2**, 402–406 (1982).
35. R Development Core Team. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2016).
36. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).

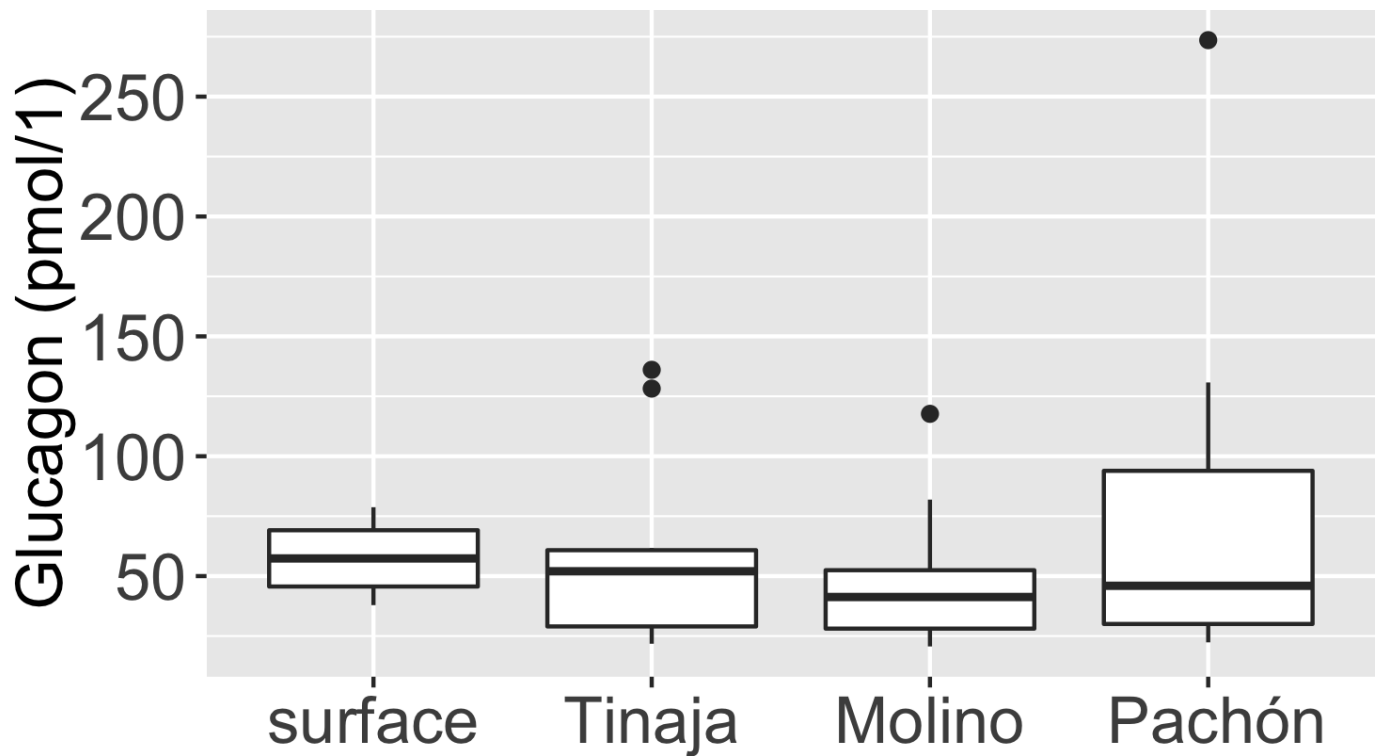


Extended Data Figure 1 | Numbers of insulin- and glucagon-positive cells in the developing pancreas are unchanged in Tinaja cavefish relative to surface fish. **a**, Whole-mount immunohistochemical detection of insulin- and glucagon-positive cells in Tinaja larvae at 10 days post fertilization. **b**, Number of glucagon- and insulin-positive cells

c

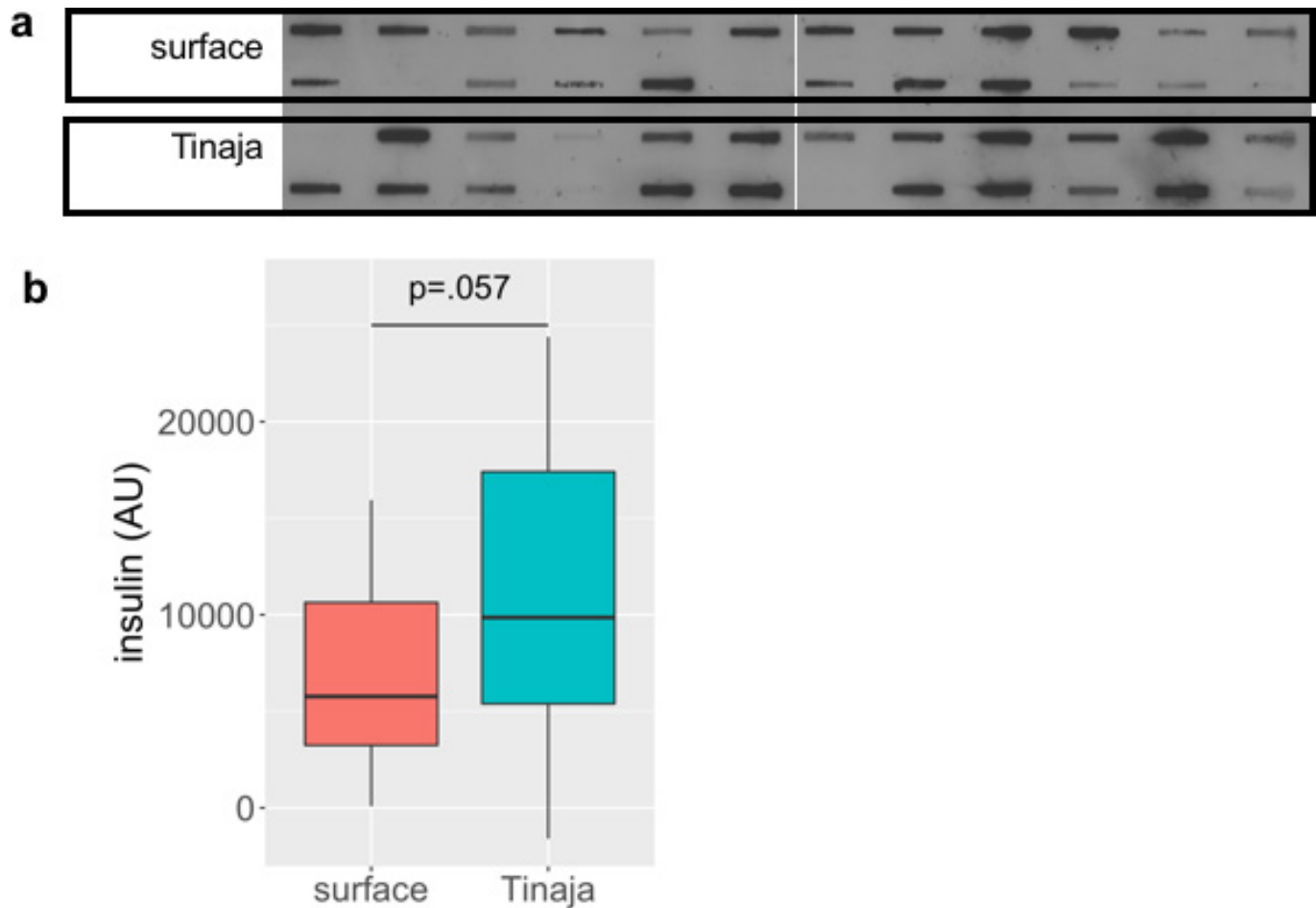
	Surface	Tinaja	p-value
<i>glucagon cells</i>	54	50	0.678
<i>insulin cells</i>	54	52	0.275
<i>fish length (mm)</i>	6.8	6.5	0.780
<i>insulin/glucagon ratio</i>	1.0	1.1	0.636

in surface and Tinaja larvae at 10–11 days post fertilization ($n = 5$ fish per population.). **c**, Average number of glucagon- and insulin-positive cells, fish length, ratio of insulin to glucagon positive cells and P value comparing the surface and Tinaja values (determined using Student's t -test).



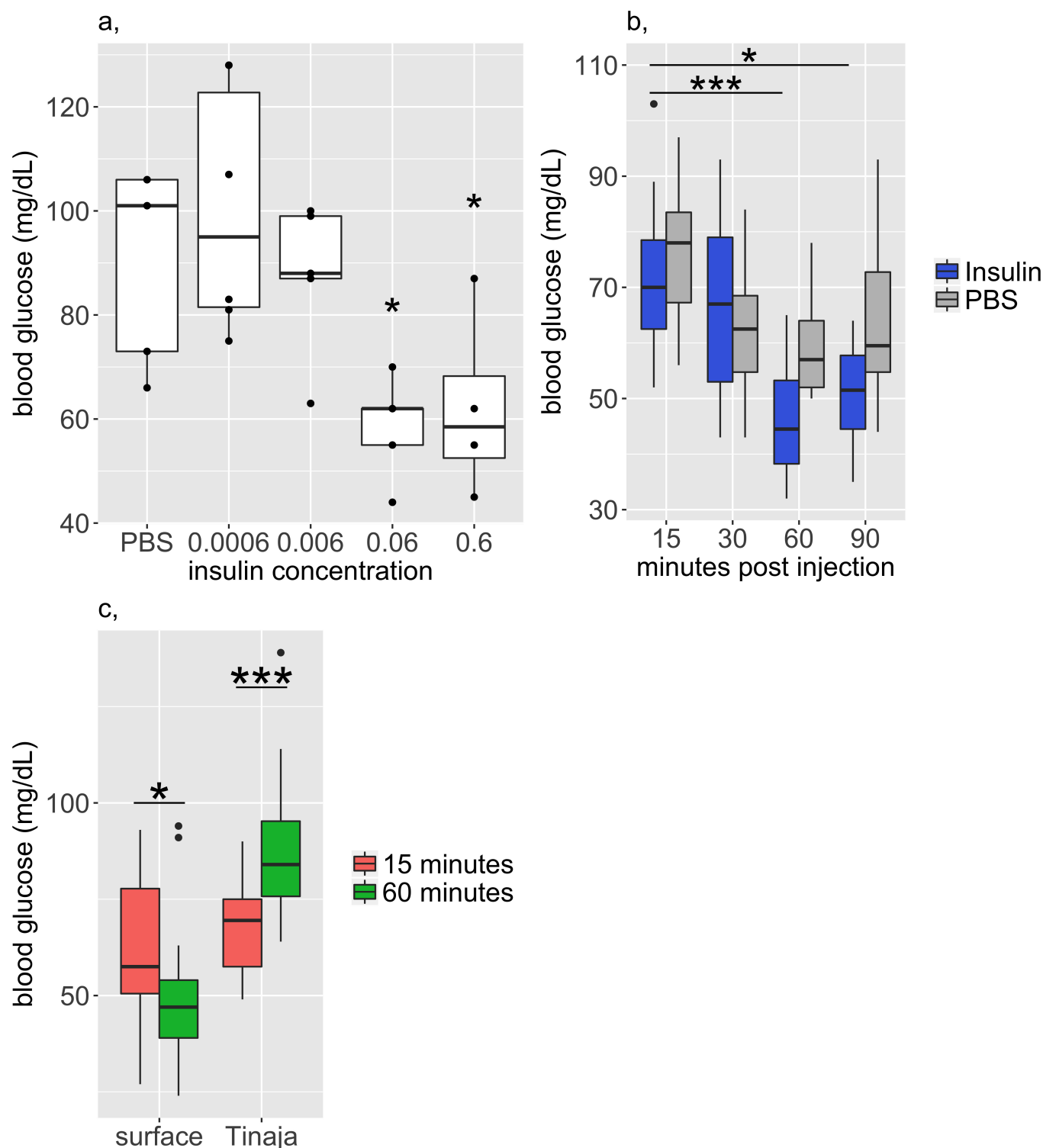
Extended Data Figure 2 | Serum glucagon levels are comparable between the different populations. Box plot comparing serum glucagon levels between surface, Tinaja, Molino and Pachón fish after 24-h fast. $n = 12$ fish per population, average of 57.87, 59.76, 79.66 and 48.89

respectively. $P = 0.52$, one-way ANOVA. Box plots show 25th, 50th and 75th percentiles (horizontal bars), and $1.5 \times$ interquartile ranges (error bars), dots represent outliers.



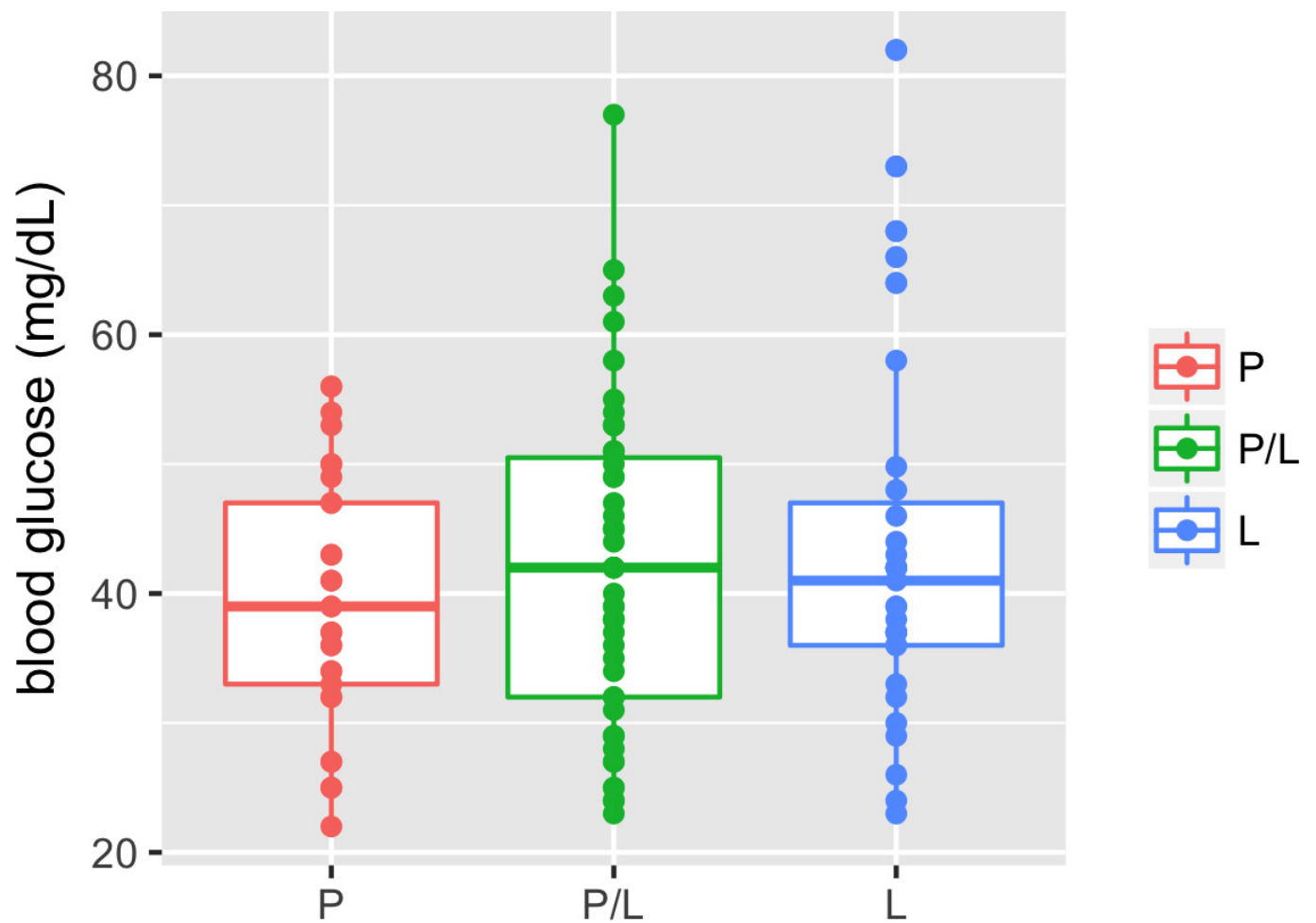
Extended Data Figure 3 | Serum insulin levels are comparable between surface and Tinaja fish. **a**, Serum blotted onto nitrocellulose membrane using Bio-Dot SF microfiltration apparatus (Bio-Rad, catalogue number 1706542) probed with anti-insulin antibody (DAKO). Each blot represents an individual fish between 1- and 2-years-old ($n = 24$ fish per population). **b**, Quantification of insulin level measured by densitometry of blots.

AU, artificial units; median, 25th, 50th and 75th percentiles (horizontal bars) and error bars at $1.5 \times$ interquartile ranges. Tinaja cavefish insulin levels (mean = 10,770) tended to be higher than those of surface fish (mean = 7,194) but the results are not significant ($P = 0.057$, Student's two-sample t -test).



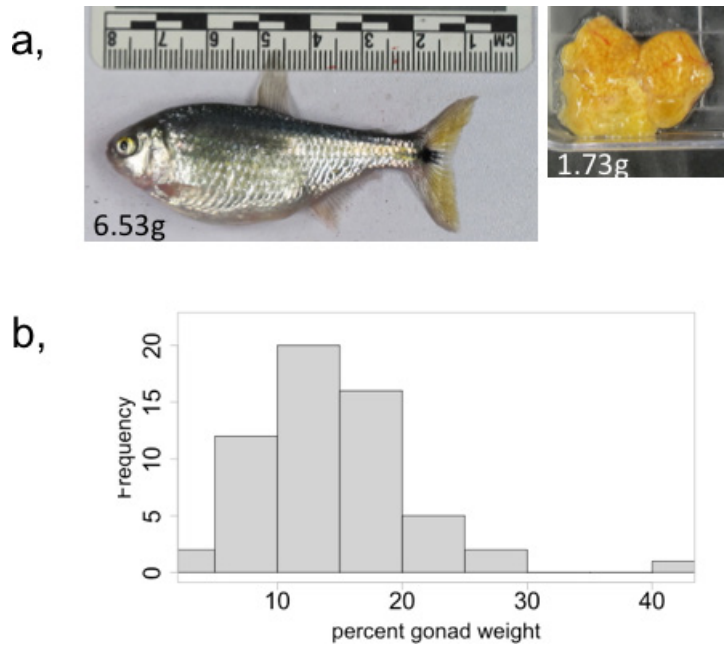
Extended Data Figure 4 | Insulin decreases blood glucose level in surface fish. **a,** We injected different concentrations of human recombinant insulin (Sigma, product I9278, stock 9.5–11.5 mg ml⁻¹) into the intraperitoneal cavity of surface fish to determine the effective dosage for subsequent experiments. Blood glucose levels are significantly lower after injection of approximately 0.6 or 0.06 g insulin per mg of fish weight compared to 0.0006 g (30 min after injection, $n = 4$ fish per dosage, dots represent individual fish, significance calculated using one-way ANOVA with Tukey's HSD post hoc test, $*P < 0.05$). We used 0.06 g insulin per mg of fish in subsequent experiments. **b,** Blood glucose levels of surface fish over time, after injection of PBS or insulin. Blood glucose levels are significantly lower at 60 and 90 min compared to 15 min after insulin injection ($n = 10$ fish per time point and condition, significance calculated using one-way ANOVA with Tukey's HSD post hoc test, $*P < 0.05$).

Therefore, we focused on the 60-min time point for comparisons with Tinaja cavefish. **c,** Blood glucose levels at 15 and 60 min after insulin injection in surface fish and Tinaja cavefish. Surface fish display a significant decrease in blood glucose levels, whereas cavefish display a significant increase in blood glucose levels (significance calculated using two-tailed Student's t -test, $*P < 0.05$, $***P < 0.0005$). Tinaja cavefish blood glucose levels may increase owing to the stress of being injected; stress hormones, such as catecholamines, ACTH and epinephrine, cause transient increases in blood glucose in humans and mice, an effect that cannot be mitigated in the absence of insulin signalling. Although both cavefish and surface fish probably undergo a stress response upon injection, this is overcome in the surface fish, which have wild-type insulin activity, but not in the Tinaja cavefish, which have reduced insulin signalling.



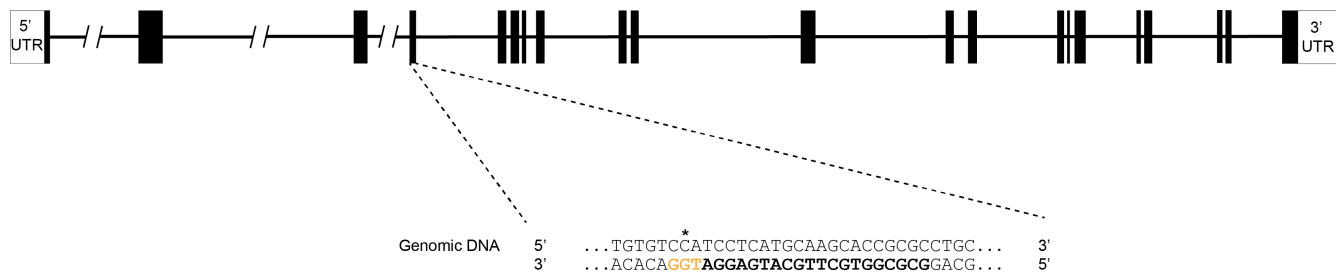
Extended Data Figure 5 | An elevated fasting blood glucose level correlates with the presence of the P211L allele in F₂ hybrids. Blood glucose levels of 192 F₂ surface-Tinaja hybrids of the indicated genotype

24 h after feeding. All of the fish with elevated blood glucose (greater than 60 mg dl⁻¹) carry the P211L allele. Median, 25th, 50th and 75th percentiles (horizontal bars), and error bars at 1.5 × interquartile ranges.



Extended Data Figure 6 | Egg mass is a confounding variable in female fish.

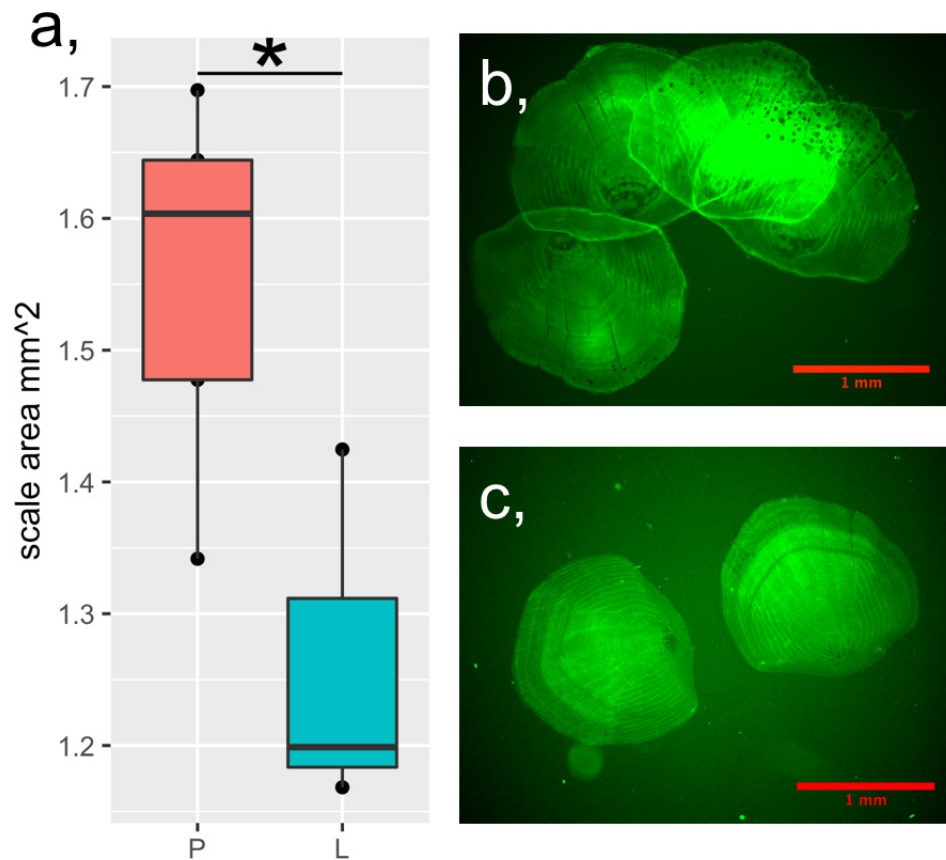
a, Image of an F₂ surface-Tinaja hybrid female and removed gonad, of indicated weights. **b,** Histogram displaying per cent gonad weight (gonad weight/total weight, multiplied by 100) of 62 F₂ surface-Tinaja hybrid females fed 6 mg per day for 4 months (minimum = 3.57, 1st quartile = 10.64, median = 13.96, mean = 13.96, 3rd quartile = 17.57 and maximum = 41.86).



ssODN 3' G*G*A*AATTTACATTGGGGACAAAGAGATAAAGAAGGGTGTACACAGATAGTAGAACATTTCGTGTCTCGGACGTGATTCTGTTGGTCACGACGGTGC*T*G*G 5'

Extended Data Figure 7 | Genome editing strategy. CRISPR–Cas9 mediated genome editing strategy in exon 3 of the *insulin receptor a* (*insra*) zebrafish gene. The guide RNA target sequence is emphasized in bold in both the reverse strand of the wild-type genomic DNA and in the ssODN.

The intended SNP exchanges are underlined, and the specific C632T to alter the P211 to L is denoted with a star. Both ends of the ssODN are protected by three phosphorothioate bonds, denoted with asterisks. The protospacer adjacent motif is shown in orange.



Extended Data Figure 8 | Scale growth is impaired in the *insra* zebrafish mutant. **a,** Quantification of scale size in zebrafish of the indicated genotype; each point represents the mean scale size of an individual fish based on the measurement of 10–14 scales removed from the left side of the body from the posterior edge of the dorsal fin to the posterior edge

of the ventral fin by gentle scraping with a scalpel. Wild type (P), $n = 4$, P211L mutant (L), $n = 3$. **b, c,** Representative images of scales stained with 0.005% calcein with contrast and brightness adjusted to show scale edges. Significance calculated using two-way t -test of mean values, $*P < 0.05$.

Reconstructing the genetic history of late Neanderthals

Mateja Hajdinjak¹, Qiaomei Fu^{2,3,4}, Alexander Hübner¹, Martin Petr¹, Fabrizio Mafessoni¹, Steffi Grote¹, Pontus Skoglund⁵, Vagheesh Narasimham⁵, Hélène Rougier⁶, Isabelle Crevecoeur⁷, Patrick Semal⁸, Marie Soressi^{9,10}, Sahra Talamo¹⁰, Jean-Jacques Hublin¹⁰, Ivan Gušić¹¹, Željko Kučan¹¹, Pavao Rudan¹¹, Liubov V. Golovanova¹², Vladimir B. Doronichev¹², Cosimo Posth^{13,14}, Johannes Krause^{13,14}, Petra Korlević¹, Sarah Nagel¹, Birgit Nickel¹, Montgomery Slatkin¹⁵, Nick Patterson^{5,16}, David Reich^{5,16,17}, Kay Prüfer¹, Matthias Meyer¹, Svante Pääbo¹⁸ & Janet Kelso¹

Although it has previously been shown that Neanderthals contributed DNA to modern humans^{1,2}, not much is known about the genetic diversity of Neanderthals or the relationship between late Neanderthal populations at the time at which their last interactions with early modern humans occurred and before they eventually disappeared. Our ability to retrieve DNA from a larger number of Neanderthal individuals has been limited by poor preservation of endogenous DNA³ and contamination of Neanderthal skeletal remains by large amounts of microbial and present-day human DNA³⁻⁵. Here we use hypochlorite treatment⁶ of as little as 9 mg of bone or tooth powder to generate between 1- and 2.7-fold genomic coverage of five Neanderthals who lived around 39,000 to 47,000 years ago (that is, late Neanderthals), thereby doubling the number of Neanderthals for which genome sequences are available. Genetic similarity among late Neanderthals is well predicted by their geographical location, and comparison to the genome of an older Neanderthal from the Caucasus^{2,7} indicates that a population turnover is likely to have occurred, either in the Caucasus or throughout Europe, towards the end of Neanderthal history. We find that the bulk of Neanderthal gene flow into early modern humans originated from one or more source populations that diverged from the Neanderthals that were studied here at least 70,000 years ago, but after they split from a previously sequenced Neanderthal from Siberia² around 150,000 years ago. Although four of the Neanderthals studied here post-date the putative arrival of early modern humans into Europe, we do not detect any recent gene flow from early modern humans in their ancestry.

The Middle to Upper Palaeolithic transition in Europe was characterized by major cultural and biological changes that coincided with the arrival of early modern humans and the disappearance of Neanderthals^{8,9}. Analyses of the first Neanderthal genomes have provided evidence of gene flow from Neanderthals into modern humans between 50,000 and 60,000 years ago (ka), resulting in around 2% of Neanderthal DNA in the genomes of non-Africans today^{1,2,10}. Additionally, genetic analyses of an approximately 39,000–42,000-year-old modern human from Romania (Oase 1) showed that interbreeding between Neanderthals and modern humans also happened in Europe at a later point in time¹¹. However, little is known about the diversity of late Neanderthal populations across Europe and western Asia shortly before their disappearance, or about their relationship to

the population that admixed with early modern humans. To date, only a handful of Neanderthal remains have been identified with a sufficiently high content of endogenous DNA and low enough levels of microbial and human DNA contamination³ to allow analysis of larger parts of their genomes^{1,2,7}, limiting our ability to study their genetic history.

In an attempt to make more Neanderthal genomes available for population analyses, we identified five specimens with sufficient preservation of endogenous DNA to explore the possibility of nuclear genome sequencing (Fig. 1a and Supplementary Information 1): a fragment of a right femur (Goyet Q56-1) dated to 43,000–42,080 calibrated years (cal. yr) before present (BP; taken to be AD 1950) from the Troisième caverne of Goyet in Belgium¹²; an upper right molar (Spy 94a) that is associated with a maxillary fragment dated to 39,150–37,880 cal. yr BP from the neighbouring Spy cave in Belgium¹³; a tooth (Les Cottés Z4-1514) dated to 43,740–42,720 cal. yr BP from Les Cottés cave¹⁴ in France; an undiagnosed bone fragment found in Vindija cave in Croatia (Vindija 87) dated to be older than 44,000 uncalibrated years BP; and a skull fragment of an infant from Mezmaiskaya cave (Mezmaiskaya 2) in the Russian Caucasus dated to 44,600–42,960 cal. yr BP¹⁵.

We extracted DNA from between 9 and 58 mg of bone or tooth powder. Approximately half of the powder from each specimen was treated with 0.5% hypochlorite solution before DNA extraction to remove present-day human and microbial DNA contamination⁶. Hypochlorite treatment increased the proportion of DNA fragments mapping to the human reference genome between 5.6- and 161-fold (Fig. 1b and Supplementary Information 2), and reduced present-day human contamination in four of the specimens between 2- and 18-fold (Fig. 1c and Supplementary Information 2). This substantial increase in the proportion of informative fragments made whole-genome sequencing of these previously inaccessible specimens feasible.

We generated additional single-stranded DNA libraries from selected extracts and sequenced them to an average genomic coverage of 2.7-fold for Les Cottés Z4-1514, 2.2-fold for Goyet Q56-1, 1.7-fold for Mezmaiskaya 2, 1.3-fold for Vindija 87 and 1-fold for Spy 94a (Extended Data Table 1 and Supplementary Information 3). Estimates of mitochondrial (mt) and nuclear DNA contamination ranged from 0.52% to 5.06% and from 0.18% to 1.75%, respectively (Supplementary Information 4). Therefore, we restricted analyses to fragments carrying cytosine (C) to thymine (T) substitutions at their ends as these derive

¹Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. ²Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. ³CAS Center for Excellence in Life and Paleoenvironment, Beijing 100044, China. ⁴University of Chinese Academy of Sciences, Beijing 100049, China. ⁵Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁶Department of Anthropology, California State University Northridge, Northridge, California 91330-8244, USA. ⁷Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France. ⁸Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium. ⁹Faculty of Archaeology, Leiden University, 2300 RA Leiden, The Netherlands. ¹⁰Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. ¹¹Croatian Academy of Sciences and Arts, HR-10000 Zagreb, Croatia. ¹²ANO Laboratory of Prehistory 14 Linia 3-11, St Petersburg 1990 34, Russia. ¹³Max Planck Institute for the Science of Human History, 07745 Jena, Germany. ¹⁴Institute for Archaeological Sciences, University of Tübingen, Rümelin Strasse 23, 72070 Tübingen, Germany. ¹⁵Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. ¹⁶Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ¹⁷Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

§These authors jointly supervised this work.

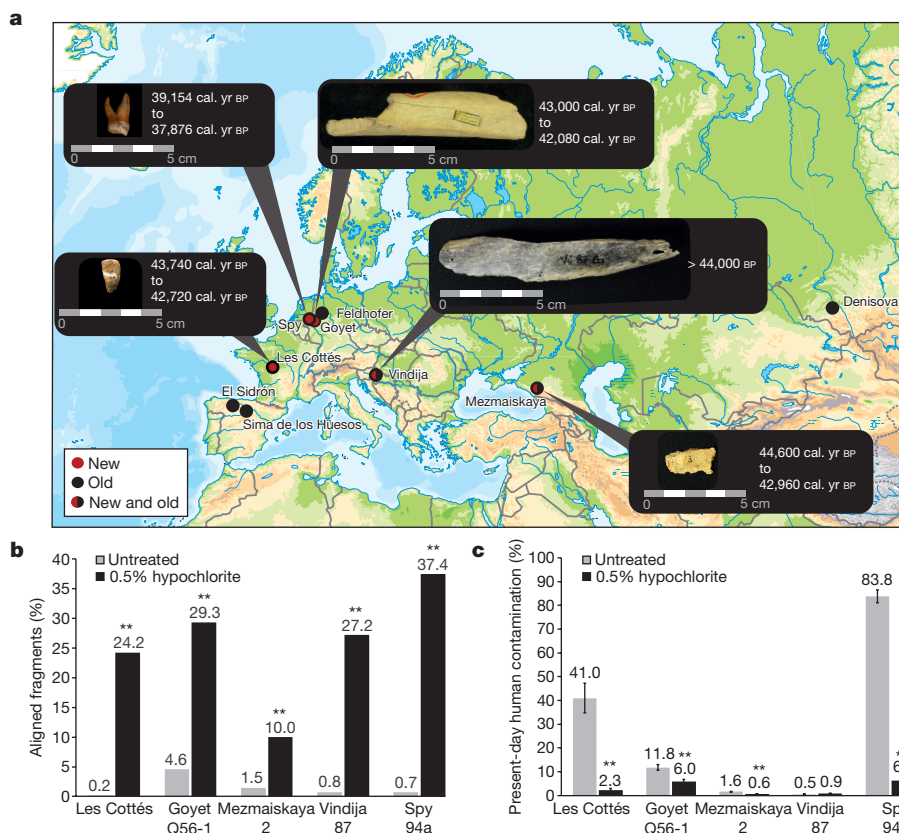


Figure 1 | Specimen information and the effects of 0.5% hypochlorite treatment. **a**, Location and age of the five late Neanderthal specimens analysed in this study (new), and other sites for which genome-wide data of Neanderthal samples have previously been published (old). Map source: Vectormapcollection. **b**, Proportion of DNA fragments aligned to the human reference genome in untreated bone and tooth powder and in

from deamination of cytosine to uracil and indicate that DNA molecules are of ancient origin^{16,17} (Extended Data Figs 1, 2). This reduced the mtDNA contamination to 0.39–1.61% and the autosomal contamination to 0–0.81% (Supplementary Information 4). To mitigate the influence of deamination on genetic inferences, we further restricted the analyses to only transversion polymorphisms (see Supplementary Information 6).

A phylogenetic tree (Fig. 2a) of the reconstructed complete mitochondrial genomes places these five specimens within the Neanderthal mtDNA variation. The relationship of the mtDNA of Les Cottés Z4-1514 to the mtDNAs of Neanderthals from across their entire geographic range—including older Neanderthals from Okladnikov and Denisova cave (Supplementary Information 5)—together with the grouping of Mezmaiskaya 2 with late Neanderthals, including Feldhofer 2, challenges the previously proposed division between Eastern and Western mtDNAs in late-surviving Neanderthals¹⁸. As inferred from the sequence coverage of the X chromosome and the autosomes, Goyet Q56-1, Les Cottés Z4-1514 and Vindija 87 were females, whereas Mezmaiskaya 2 and Spy 94a were males (Extended Data Fig. 3). The Y chromosome sequences of both male individuals fall outside the known variation of present-day human Y chromosomes (Fig. 2b and Supplementary Information 5), as is the case for the Y chromosome of a Neanderthal from El Sidrón, Spain¹⁹.

We analysed the genomes of the five late Neanderthals with the previously published high-quality genomes of an approximately 120,000-year-old Neanderthal from Siberia (Altai Neanderthal)² and Neanderthal from Croatia (Vindija 33.19) that is more than 45,000 years old⁷, the low-coverage genome of a Neanderthal who lived approximately 60–70 ka from Mezmaiskaya cave (Mezmaiskaya 1)^{2,7}, the composite low-coverage genome of three Neanderthals from

powder treated with 0.5% sodium hypochlorite (Supplementary Table 2.1). **c**, Proportion of present-day human contamination (with 95% binomial confidence intervals) inferred from mtDNA fragments in treated and untreated samples. Two-sided Pearson's χ^2 tests were used to calculate significant differences (** $\alpha \leq 0.001$) (Supplementary Table 2.5).

Croatia¹, the high-coverage genome of a Denisovan individual²⁰ and a world-wide panel of present-day humans^{2,21}. On the basis of derived alleles that are shared with the Vindija 33.19 genome⁷, we find that the Vindija 87 specimen originates from the same individual as Vindija 33.19. We therefore excluded Vindija 87 from subsequent analyses (Supplementary Information 7).

A neighbour-joining tree based on the number of pairwise transversions between individuals shows that all Neanderthals form a monophyletic clade relative to the Denisovan individual (Fig. 2c and Supplementary Information 7, 8). The tree reflects an apparent age-related division among the Neanderthals with the oldest specimen, the Altai Neanderthal branching off first, followed by Mezmaiskaya 1, whereas the late Neanderthals form a clade. To assess the relationship of the late Neanderthals to the high-coverage genomes of the Altai and Vindija 33.19 Neanderthals, we used two related statistics that measure the fraction of derived allele-sharing between the genomes. We find that all late Neanderthals and Mezmaiskaya 1 share significantly more derived alleles with Vindija 33.19 than with the Altai Neanderthal ($-32.1 \leq Z \leq -58.1$; Extended Data Table 2 and Supplementary Information 9) and that late Neanderthals share on average 49.0% (95% confidence interval, 44.2–54.2%) of the derived alleles seen in Vindija 33.19 and not in other high-coverage genomes in the analysis, whereas they share 17.2% (95% confidence interval, 15.9–18.3%) of the derived alleles in the Altai Neanderthal (Extended Data Table 3 and Supplementary Information 7).

Obtaining genome-wide data of multiple late Neanderthals from a broad geographical range enabled us to determine whether relatedness among Neanderthals is correlated to their geographical proximity, as is the case for present-day humans²². In support of this, we find that the two Neanderthals from Belgium share more derived alleles

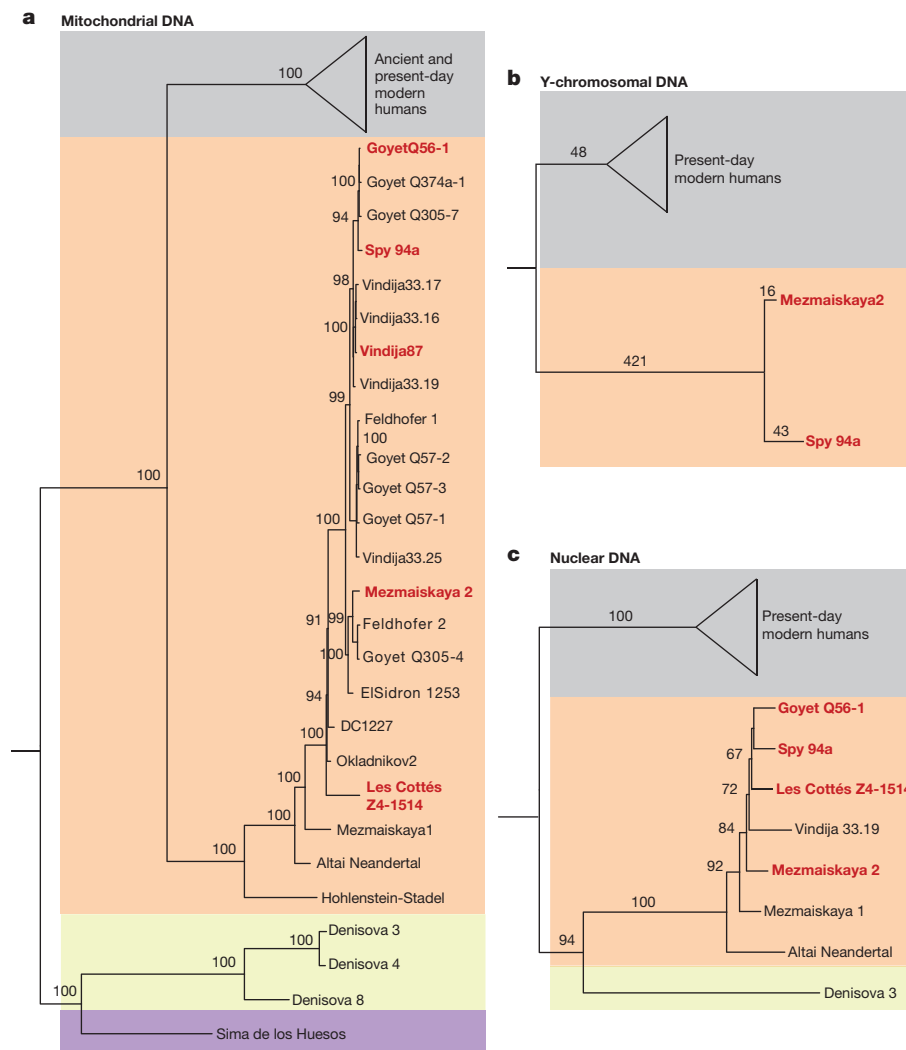


Figure 2 | Phylogenetic relationships of late Neanderthals. **a**, Bayesian phylogenetic tree of mitochondrial genomes of 23 Neanderthals, 3 Denisovans, 64 modern humans and a hominin from Sima de los Huesos. The posterior probabilities for the branches are shown. **b**, Neighbour-joining tree of Y chromosome sequences of Mezmaiskaya 2, Spy 94a, 175 present-day humans²¹ and two present-day humans

with each other than with any other Neanderthal ($-3.65 \leq Z \leq -8.47$; Supplementary Information 9), and in turn more derived alleles with Neanderthals from France and Croatia than with the late Neanderthal from the Caucasus. Similarly, the four Neanderthals from Vindija cave that come from a relatively narrow time range share more derived alleles with each other than with other Neanderthals ($-2.2 \leq Z \leq -14.5$; Supplementary Information 9). Furthermore, specimens with a similar age and the largest geographical distance between them (Les Cottés Z4-1514 and Mezmaiskaya 2) shared the fewest derived alleles (Supplementary Information 9). By contrast, Mezmaiskaya 2 shared more derived alleles with the other late Neanderthals than with Mezmaiskaya 1 ($-2.13 \leq Z \leq -9.56$; Supplementary Information 9), suggesting that there was a population turnover towards the end of Neanderthal history. This turnover may have been the result of a population related to western Neanderthals replacing earlier Neanderthals in the Caucasus, or the replacement of Neanderthals in western Europe by a population related to Mezmaiskaya 2. The timing of this turnover coincides with pronounced climatic fluctuations during Marine Isotope Stage 3 between 60 and 24 ka²³, when extreme cold periods in northern Europe may have triggered the local extinction of Neanderthal populations and subsequent re-colonization from refugia in southern Europe or western Asia^{24,25}.

carrying the A00 haplogroup³⁰. The number of substitutions is shown above the branches. **c**, Neighbour-joining tree of nuclear genomes based on autosomal transversions among late Neanderthals, Vindija 33.19, Mezmaiskaya 1, Altai Neanderthal, Denisovan and 12 present-day humans. Bootstrap support values after 1,000 replications are shown.

We estimated the population split times between each of the low-coverage Neanderthal genomes and the two high-coverage Neanderthal genomes by determining the fraction of sites at which each of the low-coverage Neanderthal genomes shares a derived allele that occurs in the heterozygote state in one of the high-quality genomes ($F(A|B)$ statistics^{1,20}). This fraction was then used to estimate the population split times for each pair of Neanderthals using previous inferences of how Neanderthal population sizes changed over time^{2,7}. Owing to the uncertainties in the mutation rate and generation times, we caution that although the times presented are likely to accurately reflect the relative ages of the population split times, the absolute estimates in years are approximate. We estimate that the late Neanderthals studied here separated from a common ancestor with the Altai Neanderthal around 150 ka (95% confidence interval, 142–186 ka), and from a common ancestor with Vindija 33.19 about 70 ka (95% confidence interval, 58–72 ka; Extended Data Table 4 and Supplementary Information 8). The estimates of the population split times from the common ancestors shared with the Denisovan and with modern humans are around 400 ka (95% confidence interval, 367–484 ka) and about 530 ka (95% confidence interval, 503–565 ka; Extended Data Table 4 and Supplementary Information 8), respectively, consistent with previous estimates using the Altai and Vindija 33.19 Neanderthal genomes^{1,2,20}.

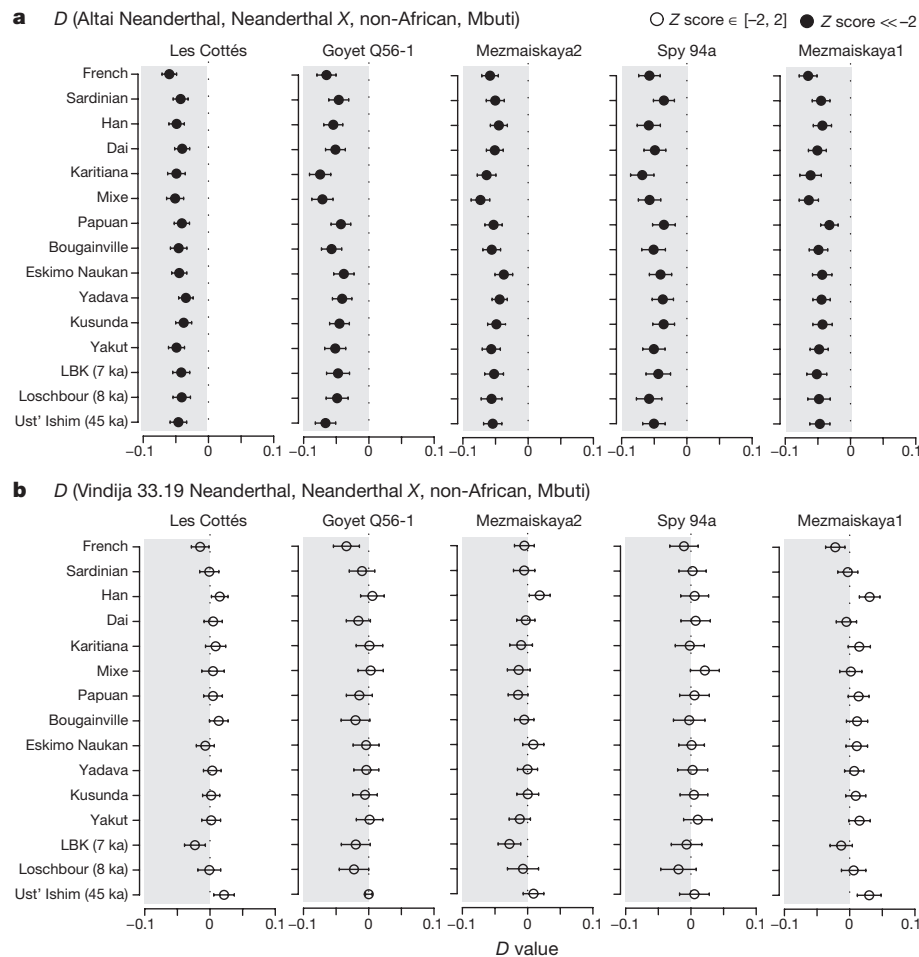


Figure 3 | Proximity to the introgressing Neanderthal populations in present-day and ancient humans calculated using D (Neanderthal1, Neanderthal2; non-African, African). Three Mbuti individuals from the Simons Genome Diversity Project (SGDP)²¹ were used as an outgroup and standard errors were calculated using a weighted block jackknife (Supplementary Information 10). Shaded grey region corresponds

to $D < 0$. **a**, All late Neanderthals and the older Mezmaiskaya 1 are significantly closer to the introgressing Neanderthal population(s) than the Altai Neanderthal. **b**, There is no significant difference between late Neanderthals, Mezmaiskaya 1 and Vindija 33.19 in their proximity to the introgressing Neanderthal population(s) in present-day and ancient humans.

To investigate whether any of the Neanderthals sequenced to date is more closely related to the Neanderthal population that contributed genetic material to modern humans, we compared the Neanderthal genomes from this and previous studies to the genomes of 263 present-day humans²¹ as well as a number of early modern humans^{10,11,26,27}. We find that all late Neanderthals and the older Mezmaiskaya 1 Neanderthal share significantly more derived alleles with the introgressing Neanderthals than the Altai Neanderthal does ($-2.4 \leq Z \leq -5.6$; Fig. 3a and Supplementary Information 10), with no significant differences among them ($-0.1 \leq Z \leq 1.8$; Fig. 3b and Supplementary Information 10). Interestingly, this is also true for an approximately 45,000-year-old modern human from Siberia (Ust' Ishim) (Fig. 3 and Supplementary Information 10), who was contemporaneous with late Neanderthals, but is not a direct ancestor of present-day humans¹⁰. Thus, the majority of gene flow into early modern humans appears to have originated from one or more Neanderthal populations that diverged from other late Neanderthals after their split from the Altai Neanderthal about 150 ka, but before the split from Mezmaiskaya 1 at least 90 ka (Extended Data Table 4). Owing to the scarcity of overlapping genetic data from Oase 1, whose genome revealed an unusually high percentage of Neanderthal ancestry¹¹, we were unable to resolve whether one of these late Neanderthals was significantly closer than others to the introgressing Neanderthal in Oase 1.

Interbreeding between Neanderthals and early modern humans is likely to have occurred intermittently, presumably resulting in gene

flow in both directions²⁸. However, when we applied an approach that uses the extended length of haplotypes expected from recent introgression into the analysed late Neanderthals, we did not find any indications of recent gene flow from early modern humans to the late Neanderthals (Supplementary Information 11). We caution that given the small number of analysed Neanderthals we cannot exclude that such gene flow occurred. However, it is striking that Oase 1, one of two early modern humans that overlapped in time with late Neanderthals, showed evidence for recent additional Neanderthal introgression^{10,11} whereas none of the late Neanderthals analysed here do. This may indicate that gene flow affected the ancestry of modern human populations more than it did Neanderthals²⁹. Further work is necessary to determine whether this was the case. Our work demonstrates that the generation of genome sequences from a large number of archaic human individuals is now technically feasible, and opens up the possibility to study Neanderthal populations across their temporal and geographical range.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 July 2017; accepted 24 January 2018.

Published online 21 March 2018.

- Green, R. E. *et al.* A draft sequence of the Neanderthal genome. *Science* **328**, 710–722 (2010).

2. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
3. Pääbo, S. *et al.* Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
4. Gilbert, M. T., Bandelt, H. J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005).
5. Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236 (2010).
6. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
7. Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).
8. Hublin, J.-J. The modern human colonization of western Eurasia: when and where? *Quat. Sci. Rev.* **118**, 194–210 (2015).
9. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306–309 (2014).
10. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
11. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
12. Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci. Rep.* **6**, 29005 (2016).
13. Semal, P. *et al.* New data on the late Neandertals: direct dating of the Belgian Spy fossils. *Am. J. Phys. Anthropol.* **138**, 421–428 (2009).
14. Soressi, M. *et al.* in *Préhistoire entre Vienne et Charente — Hommes et sociétés du Paléolithique* mémoire 38 (eds Buisson-Catil, J. & Primault, J.) 221–234 (Association des Publications Chauvinoises, 2010).
15. Pinhasi, R., Higham, T. F., Golovanova, L. V. & Doronichev, V. B. Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc. Natl Acad. Sci. USA* **108**, 8611–8616 (2011).
16. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
17. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406 (2014).
18. Dalén, L. *et al.* Partial genetic turnover in Neandertals: continuity in the East and population replacement in the West. *Mol. Biol. Evol.* **29**, 1893–1897 (2012).
19. Mendez, F. L., Poznik, G. D., Castellano, S. & Bustamante, C. D. The divergence of Neandertal and modern human Y chromosomes. *Am. J. Hum. Genet.* **98**, 728–734 (2016).
20. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
21. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
22. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, 1994).
23. Svensson, A. *et al.* A 60 000 year Greenland stratigraphic ice core chronology. *Clim. Past* **4**, 47–57 (2008).
24. Hublin, J.-J. & Roebroeks, W. Ebb and flow or regional extinctions? On the character of Neandertal occupation of northern environments. *C. R. Palevol* **8**, 503–509 (2009).
25. Müller, U. C. *et al.* The role of climate in the spread of modern humans into Europe. *Quat. Sci. Rev.* **30**, 273–279 (2011).
26. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
27. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
28. Kuhlwilm, M. *et al.* Ancient gene flow from early modern humans into Eastern Neandertals. *Nature* **530**, 429–433 (2016).
29. Overmann, K. & Coolidge, F. Human species and mating systems: Neandertal–Homo sapiens reproductive isolation and the archaeological and fossil records. *J. Anthropol. Sci.* **91**, 91–110 (2013).
30. Karmin, M. *et al.* A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **25**, 459–466 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Weihmann and B. Höber for their help with DNA sequencing, U. Stenzel for computational support and advice for data analysis, R. Barr for the help with the graphics, V. Slon for helpful discussions and comments on the manuscript. Q.F. is funded in part by NSFC (91731303, 41672021, 41630102), CAS (QYZDB-SSW-DQC003, XDB13000000, XDA19050102, XDPB05) and the Howard Hughes Medical Institute (grant number 55008731). D.R. is supported by the US National Science Foundation (grant BCS-1032255) and by an Allen Discovery Center of the Paul Allen Foundation and is an investigator of the Howard Hughes Medical Institute. This study was funded by the Max Planck Society and the European Research Council (grant agreement number 694707 to S.P.). M.So. thanks the owner of Les Cottés, and the French Ministry of Culture for financial support and excavation permits.

Author Contributions M.H., M.M. and S.P. conceived the study. M.Sl., N.P., D.R., K.P., M.M., S.P. and J.Ke. supervised the study. M.H., P.K., S.N. and B.N. performed ancient DNA laboratory work. H.R., I.C., P.Se., M.So., S.T., J.-J.H., I.G., Ž.K., P.R., L.V.G., V.B.D., C.P. and J.Kr. provided and analysed archaeological material. M.H., Q.F., A.H., M.P., F.M., S.G., P.Sk. and V.N. analysed ancient DNA data. M.H., M.M., S.P. and J.Ke. wrote the manuscript with the input of all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.H. (mateja_hajdinjak@eva.mpg.de), S.P. (paabo@eva.mpg.de) or J.K. (kelso@eva.mpg.de).

Reviewer Information *Nature* thanks C. Lalueza-Fox, C. Stringer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

DNA extraction and library preparation. All specimens were sampled in clean room facilities dedicated to the analysis of ancient DNA. Between 28 mg and 104 mg of tooth or bone powder was obtained by drilling once into the physically cleaned part of the specimen and split evenly (Supplementary Information 2; Supplementary Table 2.1). Approximately half of the powder was directly subjected to DNA extraction using a silica-based method³¹ as implemented in ref. 6 (the untreated sample), whereas the second half was treated with a 0.5% sodium hypochlorite solution⁶ before DNA extraction in an attempt to remove some of the microbial and present-day human DNA contamination^{3–5}. Subsequently, 5 or 10 µl of each extract was converted into single-stranded DNA libraries³² with the modifications as in ref. 6. The initial libraries of Vindija 87, Goyet Q56-1 and Les Cottés Z4-1514 were treated with *Escherichia coli* uracil-DNA-glycosylase (UDG) and *E. coli* endonuclease VIII (Endo VIII)^{15,20} to excise uracils, while all other libraries were prepared without this enzymatic treatment (Supplementary Information 2). The libraries were amplified into plateau³³ and tagged with two sample-specific indices^{6,34}. An aliquot of each amplified library was additionally enriched for hominin mtDNA using a bead-based hybridization method and modern human mtDNA as a bait^{35–37}. After analysing the first set of libraries, we selected the extracts with the highest proportion of endogenous DNA and the lowest levels of present-day human DNA contamination to produce the final set of single-stranded DNA libraries^{6,32} (Supplementary Information 2 and Supplementary Table 2.6).

Genome sequencing and data processing. All libraries were initially sequenced together on Illumina's MiSeq and HiSeq 2500 platforms to determine their suitability for whole-genome sequencing. Subsequently, 23 libraries from five Neanderthal specimens were selected and sequenced on 50 lanes of the Illumina HiSeq 2500 platform in rapid mode, using double index configuration (2 × 76 bp)³⁴ (Supplementary Information 2). Base calling was done using Bustard (Illumina) for the MiSeq runs and FreeBis³⁸ for the HiSeq runs. Adapters were trimmed and overlapping paired-end reads were merged into single sequences using leeHom³⁹. The Burrows–Wheeler aligner (BWA, version 0.5.10-evan.9-1-g44db244; <https://github.com/mpieva/network-aware-bwa>)⁴⁰ was used to align the shotgun data to the modified human reference GRCh37 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/) and to align the human mtDNA capture data to the revised Cambridge Reference Sequence (NC_01290) with parameters adjusted for ancient DNA ($-n\ 0.01 -o\ 2 -l\ 16500$)²⁰. We developed a two-step algorithm called jivebunny (<https://bioinf.eva.mpg.de/jivebunny>) for de-multiplexing of the sequencing runs (Supplementary Information 3), retaining only fragments that were assigned to the correct library based on their index sequences for all of the downstream analyses. PCR duplicates were removed using bam-rmdup (version 0.6.3; <https://github.com/mpieva/biohazard-tools/>) and fragments were filtered for read length (≥ 35 bp) and mapping quality ($MQ \geq 25$) using SAMtools (version 1.3.1)⁴¹.

The deamination of cytosine (C) to uracil (U) residues leaves characteristic C-to-T substitution patterns in ancient DNA molecules, which are particularly close to the alignment ends¹⁵ and thus provide evidence for the presence of authentic ancient DNA in specimens^{5,42} (Supplementary Information 2, 3 and Extended Data Fig. 1). We evaluated the frequency of these nucleotide substitution patterns characteristic of ancient DNA using an in-house Perl script. For determining the coverage of the nuclear genomes, we counted the number of bases with a base quality of at least 30 ($BQ \geq 30$) in the fragments that overlapped highly mappable regions of the autosomes of the human genome (Map35_100% of ref. 2) and divided this number by the total length of those regions. We determined the sex of the Neanderthal individuals by counting the number of fragments that aligned to the X chromosome and the autosomes (Extended Data Fig. 3).

Contamination estimates. We used four methods to estimate the proportion of present-day human DNA contamination in the final dataset (Supplementary Information 4). We estimated the proportion of mtDNA contamination by present-day human DNA using two different sets of positions. First, we re-aligned the shotgun data of all HiSeq runs to the revised Cambridge Reference Sequence (rCRS, NC_012920) using BWA⁴⁰. We then counted the number of fragments overlapping 63 positions at which 18 published Neanderthal mitochondrial genomes^{1,2,12,43–46} differ from those of 311 present-day humans⁴³. In the second approach, we determined positions in the reconstructed mtDNA genomes of each of the five Neanderthals that are specific for that Neanderthal compared to 311 present-day humans. To mitigate the effect of deamination for both approaches, we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G¹⁷.

We estimated the extent of present-day human DNA contamination in the autosomal sequences of the five late Neanderthals using a maximum likelihood

approach described in ref. 1, based on all sites covering informative positions in the nuclear genome at which humans carry a fixed derived variant when compared to the great apes. Additionally, we estimated levels of present-day human DNA contamination using an ancestry model in which each low-coverage Neanderthal traces a portion of the genome either from a high-coverage uncontaminated Neanderthal, or from a modern human population related to present-day non-Africans. We built a two-source qpAdm⁴⁷ model in which one part of the ancestry was modelled as being closely related to the high-coverage genome of Vindija 33.19 Neanderthal, and the other source of ancestry was modelled as being most closely related to the Dinka population of the Simons Genome Diversity Project (SGDP)²¹. We estimate the proportion of male contamination for the three female individuals, Les Cottés Z4-1514, Goyet Q56-1 and Vindija 87 by counting the number of fragments aligning to the unique regions of the Y chromosome and divided it with the number of fragments that would be expected if the individual was a male (Supplementary Information 4 and Supplementary Table 4.1).

mtDNA and Y chromosomes of Neanderthals. We re-aligned the shotgun data of all HiSeq runs to the Vindija 33.16 mitochondrial genome (AM948965)⁴³ using BWA⁴⁰ to reconstruct complete mitochondrial genomes of five Neanderthals. A consensus base was called at each position covered by at least three fragments and where at least two-thirds of fragments had an identical base¹⁷ (Supplementary Information 5). We aligned the reconstructed mitochondrial genomes of the five late Neanderthals to the mtDNA genomes of 18 Neanderthals^{2,12,43–46,48,49}, 311 present-day humans⁴³, 10 ancient modern humans^{5,10,36,50–52}, 3 Denisovans^{53–55}, a hominin from Sima de los Huesos¹⁷ and a chimpanzee⁵⁶ using MAFFT⁵⁷ and calculated the number of pairwise differences among mitochondrial genomes using MEGA⁵⁸. Bayesian phylogenetic analysis was performed using BEAST version 2.4.5⁵⁹ to infer the time to the most recent common ancestor of all Neanderthal mitochondrial genomes (Supplementary Information 5). Radiocarbon dates of Neanderthals and ancient modern humans were used as calibration points for the molecular clock and we used jModelTest⁶⁰ and marginal likelihood estimation analysis^{61,62} to choose the best fitting substitution, clock and tree model (Supplementary Information 5).

For reconstructing the phylogeny of the Neanderthal Y chromosomes, we compared the two male Neanderthal individuals, Mezmaiskaya 2 and Spy 94a, to 175 present-day human males from the SGDP²¹ and two present-day humans with haplogroup A00³⁰. For processing the Y chromosomal data of the presented Neanderthal individuals, we followed the processing of the previously published Neanderthal Y chromosome analysis¹⁹ and used the described parameters to call bases and infer genotypes (Supplementary Information 5).

Set of filters for nuclear data analyses. As the data presented in this study were of low-coverage and contained elevated C-to-T substitutions, we investigated spurious correlations stemming from the properties of the data using D -statistics^{1,54,63} (Supplementary Information 6). All of the sampling schemes, except random read sampling and then restricting to transversion polymorphisms, as well as simulation of UDG treatment, resulted in significant spurious correlations between individuals (Supplementary Information 6 and Supplementary Table 6.1). Therefore, we applied random read sampling at each position in the genome that was covered for the low-coverage individuals. To diminish the effect of present-day human DNA contamination and enrich for the endogenous fragments¹⁷, we further selected the fragments that showed C-to-T substitutions relative to the human reference genome at the first three and/or the last three positions, that is, putatively deaminated fragments. The newly generated sequencing data of Mezmaiskaya 1⁷, an approximately 60,000–70,000-year-old sample of a Neanderthal from Russia, were processed in the same way as the data of the other low-coverage individuals. For the high-coverage genomes of Altai² and Vindija 33.19⁷ Neanderthals, as well as the Denisovan²⁰ individual, snpAD genotype calls (<http://cdna.eva.mpg.de/neanderthal/Vindija/VCF/>) were used. For comparison to modern humans, we used hetfa files of 263 present-day human genomes of the SGDP²¹ and snpAD genotype calls of three high quality ancient modern humans, Ust'-Ishim, an approximately 45,000-year-old modern human from Siberia¹⁰; Loschbour, an around 8,000-year-old hunter-gatherer from Luxembourg²⁶; and LBK, an approximately 7,000-year-old farmer from Stuttgart²⁶. Variant sites across all genomes were extracted (<https://bioinf.eva.mpg.de/heffalump>) and converted into an input format for AdmixTools (version 4.1)⁶³ or exported into combined VCF files. We further restricted all analyses to bi-allelic sites in the genome covered by at least one low-coverage Neanderthal genome and to transversion polymorphisms.

Principal component analysis. We carried out principal component analysis^{64,65} using genomes of Vindija 33.19⁷, the Altai Neanderthal² and the Denisovan individual²⁰ to estimate the eigenvectors of the genetic variation and then projected the low-coverage Neanderthal genomes onto the defined plane. This allowed us to explore the relationship of low-coverage Neanderthal genomes relative to the high-coverage Neanderthal genomes (Extended Data Fig. 4).

Lineage attribution and average sequence divergence of late Neanderthals. We followed an approach described in ref. 66 based on the sharing of derived alleles with a certain hominin group to determine more precisely to which hominin group the nuclear genomes of Les Cottés Z4-1514, Goyet Q56-1, Spy 94a, Mezmaiskaya 2 and Vindija 87 are most closely related. We investigated the state of DNA fragments overlapping the positions at which the high-coverage genomes of the Altai Neanderthal², the Vindija 33.19 Neanderthal⁷, the Denisovan individual²⁰ and a present-day African (Mbuti, HGDP00982)² differ from those of the great apes (chimpanzee, bonobo, gorilla and orangutan). We then determined the proportion of fragments for each of the low-coverage Neanderthal genomes that supported the derived state of each of the branches in the phylogenetic tree relating the four high-coverage genomes⁶⁶ (Extended Data Table 3, Supplementary Information 7 and Supplementary Table 7.1).

We estimated the divergence of the five late Neanderthal genomes along the lineage from the ancestor shared with the chimpanzee and the high-coverage genomes of the Altai and Vindija 33.19 Neanderthals, the Denisovan individual, or a present-day human from the B-panel of ref. 2, using the triangulation method previously applied to a number of ancient genomes^{1,2,20,54,55}. We calculated how many of the substitutions that were inferred to have occurred from the human–chimpanzee ancestor to the high-coverage genomes that occurred after the split from the low-coverage genome (Supplementary Information 7 and Supplementary Tables 7.2, S7.3). Standard errors were computed by a weighted block jackknife⁶⁷ with a block size of 5 million base pairs (5 Mb) across all autosomes.

Split times of late Neanderthals and the neighbour-joining tree of nuclear genomes. We conditioned on the heterozygous sites in the two high-coverage Neanderthals, and then computed the fraction of sites that show the same derived allele in randomly sampled fragments of the low-coverage individuals in order to estimate the split times between the low-coverage Neanderthals and the Altai² and Vindija 33.19⁷ Neanderthals (Extended Data Table 4 and Supplementary Information 8). In this $F(A|B)$ statistic^{1,2,20}, the expected value depends only on the demography of the high-coverage individuals from which heterozygous sites are determined. We performed these analyses on all randomly sampled fragments, and on deaminated fragments only (Extended Data Table 4 and Supplementary Information 8).

We used the low-coverage nuclear genomes of Les Cottés Z4-1514, Goyet Q56-1, Spy 94a, Mezmaiskaya 2 and Mezmaiskaya 1, the high-coverage genomes of Vindija 33.19, Altai Neanderthal, the Denisovan individual and 12 present-day humans from ref. 2 for the construction of a neighbour-joining tree. We counted the total number of transversions between all pairs of individuals and the human–chimpanzee common ancestor^{2,54}. We constructed a neighbour-joining tree⁶⁸ based on the pairwise number of transversions in windows of 5 Mb across all autosomes between all individuals with 1,000 bootstrap replications (Supplementary Information 8). The tree was constructed as implemented in the R-package phangorn⁶⁹ and visualized with FigTree (version 1.4.2) (<http://tree.bio.ed.ac.uk/software/figtree/>).

Inferring the relationships between Neanderthals. We used D -statistics^{1,54,63} to investigate the population relationships among Neanderthal individuals. We co-analysed all low-coverage Neanderthal genomes with the high-coverage genomes of the Altai² and Vindija 33.19⁷ Neanderthals. We used the whole-genome alignments of the chimpanzee, orangutan and rhesus macaque to the human reference genome^{70–72} to infer the ancestral states for the analyses of D (Altai, Vindija 33.19; Neanderthal, outgroup). Furthermore, we used the genomes of the Dinka and Mbuti individuals from the SGDP²¹ as outgroups for the statistics of D (Neanderthal1, Neanderthal2; Neanderthal3, outgroup). The standard errors were computed using a weighted block jackknife^{63,67} with equally sized blocks of 5 Mb over all autosomes. We further restricted these analyses to bi-allelic sites in the genome covered by at least one low-coverage Neanderthal and transversion polymorphisms.

Inferring the relationship to the introgressed Neanderthals in present-day and ancient modern humans. We analysed the low-coverage late Neanderthal genomes together with the high-coverage genomes of the Altai² and Vindija 33.19 Neanderthals⁷, the high-coverage genomes of the Denisovan individual²⁰ and 263 present-day humans of the SGDP²¹ (Supplementary Information 10). We included Ust'-Ishim, an approximately 45,000-year-old modern human from Siberia¹⁰; Loschbour, an around 8,000-year-old hunter-gatherer from Luxembourg²⁶; and LBK, an around 7,000-year-old farmer from Stuttgart²⁶ to study the differences among Neanderthals in their proximity to the introgressed Neanderthal DNA detected in ancient modern humans. Analyses were restricted to transversion polymorphisms and to bi-allelic sites in the genome covered by at least one low-coverage Neanderthal. As above, D -statistics were used to infer the relationships between individuals^{1,54,63} and standard errors were computed using a weighted block jackknife^{63,67} over all autosomes (block size: 5 Mb).

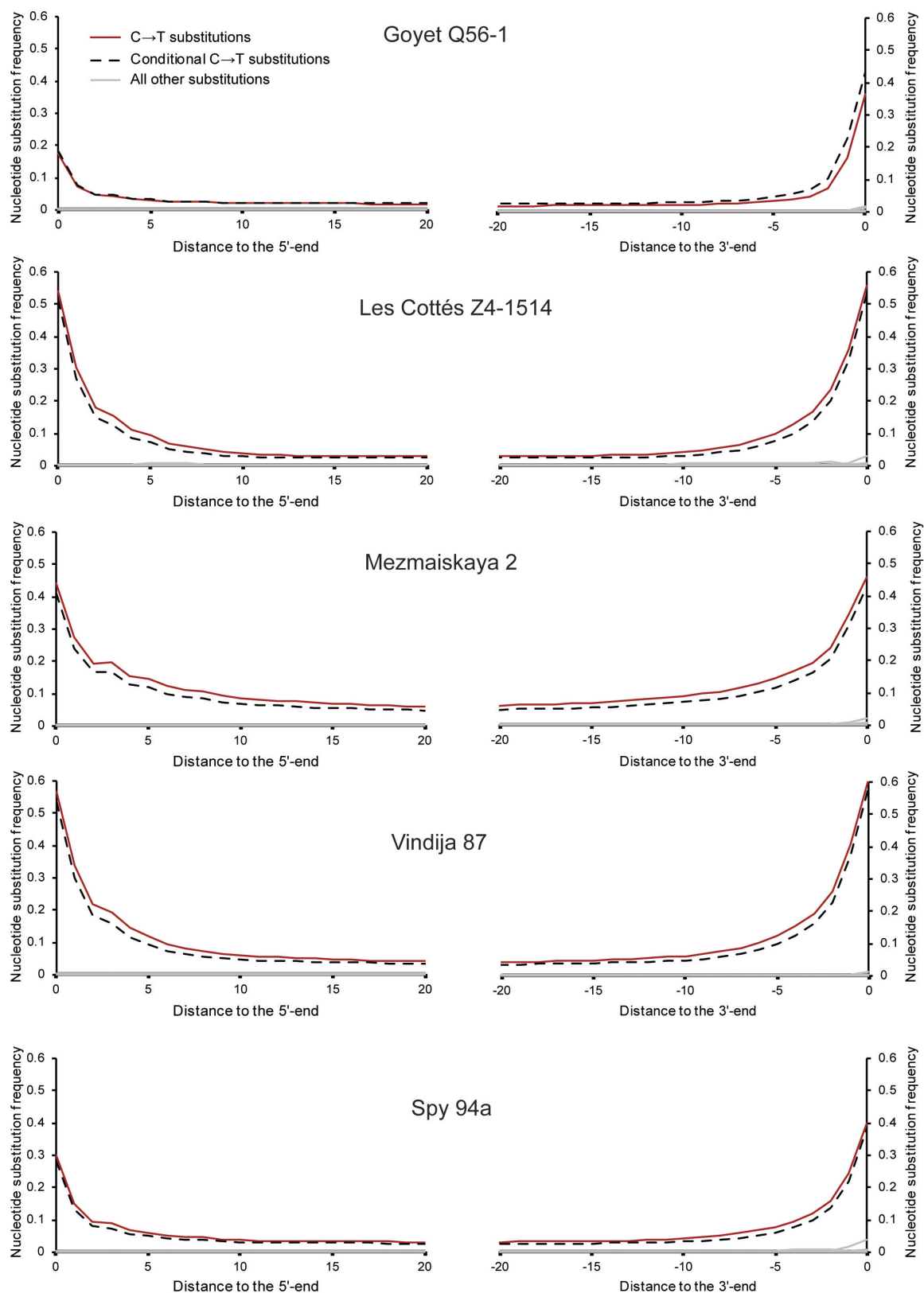
Early modern human gene flow into late Neanderthals. We modelled admixture from modern humans into Neanderthals with an ascertainment scheme in which both the Denisovan individual and Altai Neanderthal were fixed for the ancestral allele and at least half of the alleles in present-day African populations are derived. We applied the method as previously described⁷³ and estimated the date of recent early modern human admixture into Neanderthals to be around 10–100 generations ago. At each single-nucleotide polymorphism in the genome, we considered data from all Yoruba individuals from the 1000 Genomes Project⁷⁴ covered by at least three fragments that passed a pre-defined set of filters. Furthermore, we restricted the analysis to sites in the genome at which ≥ 24 Yoruba individuals as well as the Altai Neanderthal and Denisovan individual had allele calls (Map35_50% filter from ref. 2). The ancestral states were taken from the inferred ancestor of humans and chimpanzees (Ensembl Compara version 64)^{75,76}. We introduced a more complex demographic history that is loosely based on the model described in ref. 77 (details in Supplementary Information 11).

Code availability. All software packages used for analysis are cited and are publicly available. All custom developed software is freely available at <https://bioinf.eva.mpg.de/>.

Data availability. The aligned sequences have been deposited in the European Nucleotide Archive under accession numbers PRJEB21870 (Goyet Q56-1), PRJEB21875 (Les Cottés Z4-1514), PRJEB21881 (Mezmaiskaya 2), PRJEB21882 (Vindija 87) and PRJEB21883 (Spy 94a). The mitochondrial consensus sequences reported in this paper are available in GenBank with accession numbers MG025536, MG025537, MG025538, MG025539 and MG025540.

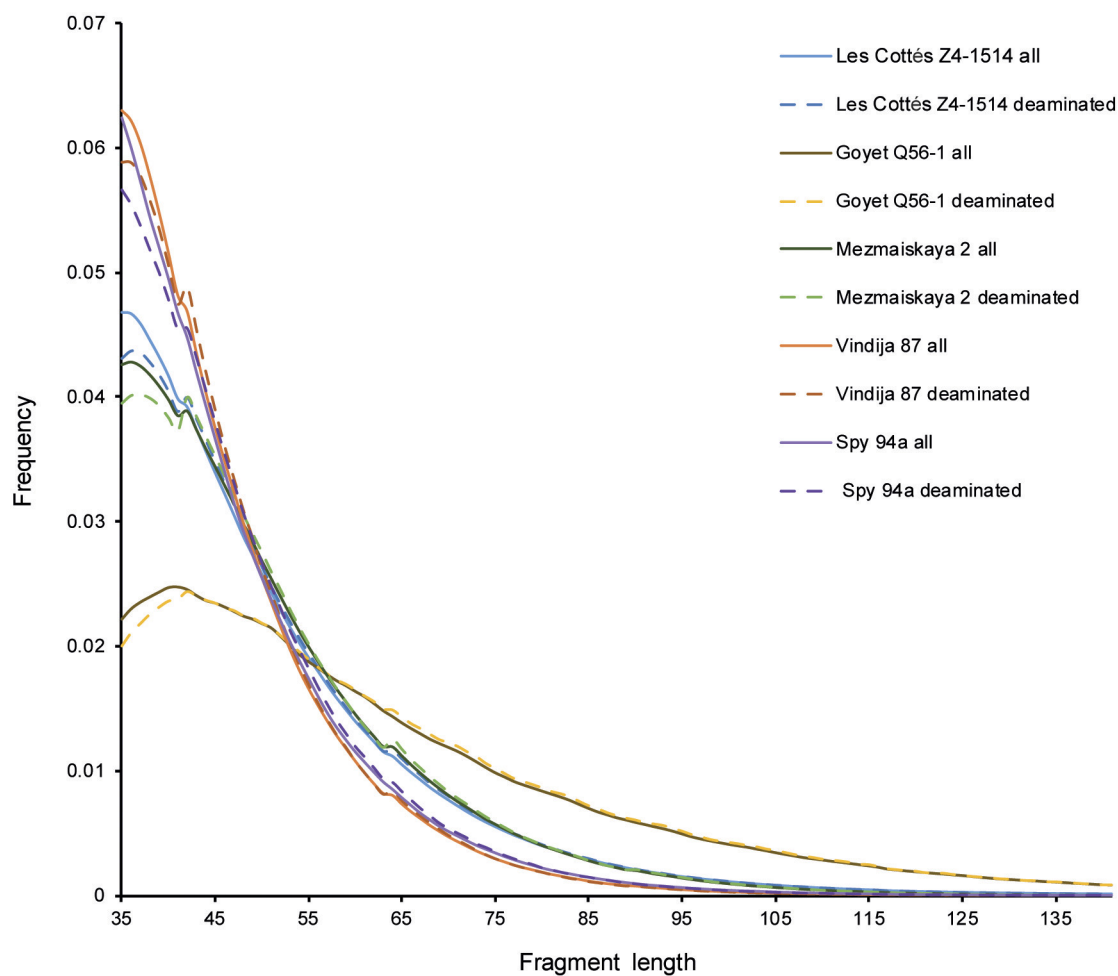
- Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
- Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
- Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012).
- Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
- Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).
- Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
- Welker, F. *et al.* Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl Acad. Sci. USA* **113**, 11162–11167 (2016).
- Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freeBis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208–1209 (2013).
- Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
- Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131 (2012).
- Green, R. E. *et al.* A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
- Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neanderthal mtDNA genomes. *Science* **325**, 318–321 (2009).
- Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neanderthal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
- Brown, S. *et al.* Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6**, 23559 (2016).
- Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
- Gansauge, M. T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24**, 1543–1549 (2014).
- Posth, C. *et al.* Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat. Commun.* **8**, 16046 (2017).
- Ermini, L. *et al.* Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr. Biol.* **18**, 1687–1693 (2008).
- Gilbert, M. T. P. *et al.* Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787–1789 (2008).
- Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897 (2010).

54. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
55. Sawyer, S. *et al.* Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl Acad. Sci. USA* **112**, 15696–15700 (2015).
56. Horai, S. *et al.* Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35**, 32–43 (1992).
57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
58. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
59. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
60. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
61. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).
62. Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A. & Lemey, P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol. Biol. Evol.* **30**, 239–243 (2013).
63. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
64. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
65. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
66. Meyer, M. *et al.* Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).
67. Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete-m jackknife for unequal m. *Stat. Comput.* **9**, 3–8 (1999).
68. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
69. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
70. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
71. Locke, D. P. *et al.* Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529–533 (2011).
72. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
73. Moorjani, P. *et al.* A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl Acad. Sci. USA* **113**, 5652–5657 (2016).
74. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
75. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* **18**, 1814–1828 (2008).
76. Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18**, 1829–1843 (2008).
77. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl Acad. Sci. USA* **108**, 11983–11988 (2011).



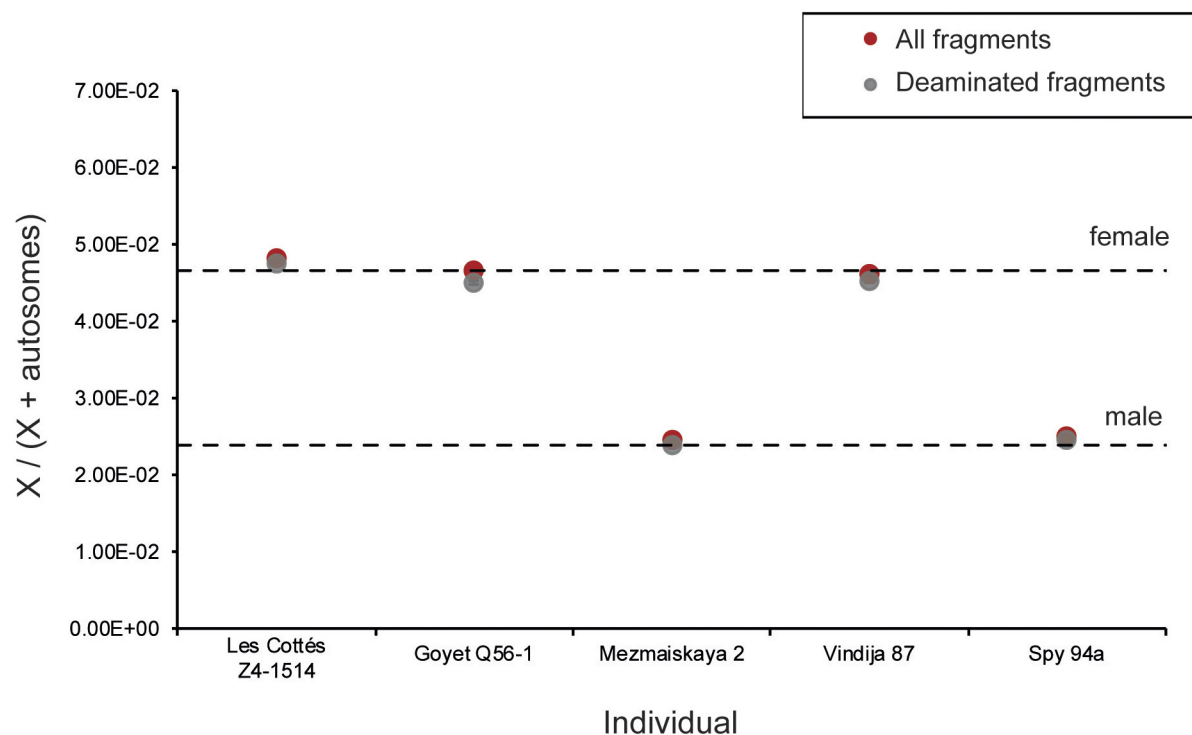
Extended Data Figure 1 | Frequency of nucleotide substitutions at the beginning and the end of nuclear alignments for the final dataset of Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87 and Spy 94a. Only fragments of at least 35 bp that mapped to the human reference

genome with a mapping quality of at least 25 ($MQ \geq 25$) were used for this analysis. Solid lines depict all fragments and dashed lines the fragments that have a C-to-T substitution at the opposing end ('conditional' C-to-T substitutions). All other types of substitutions are marked in grey.



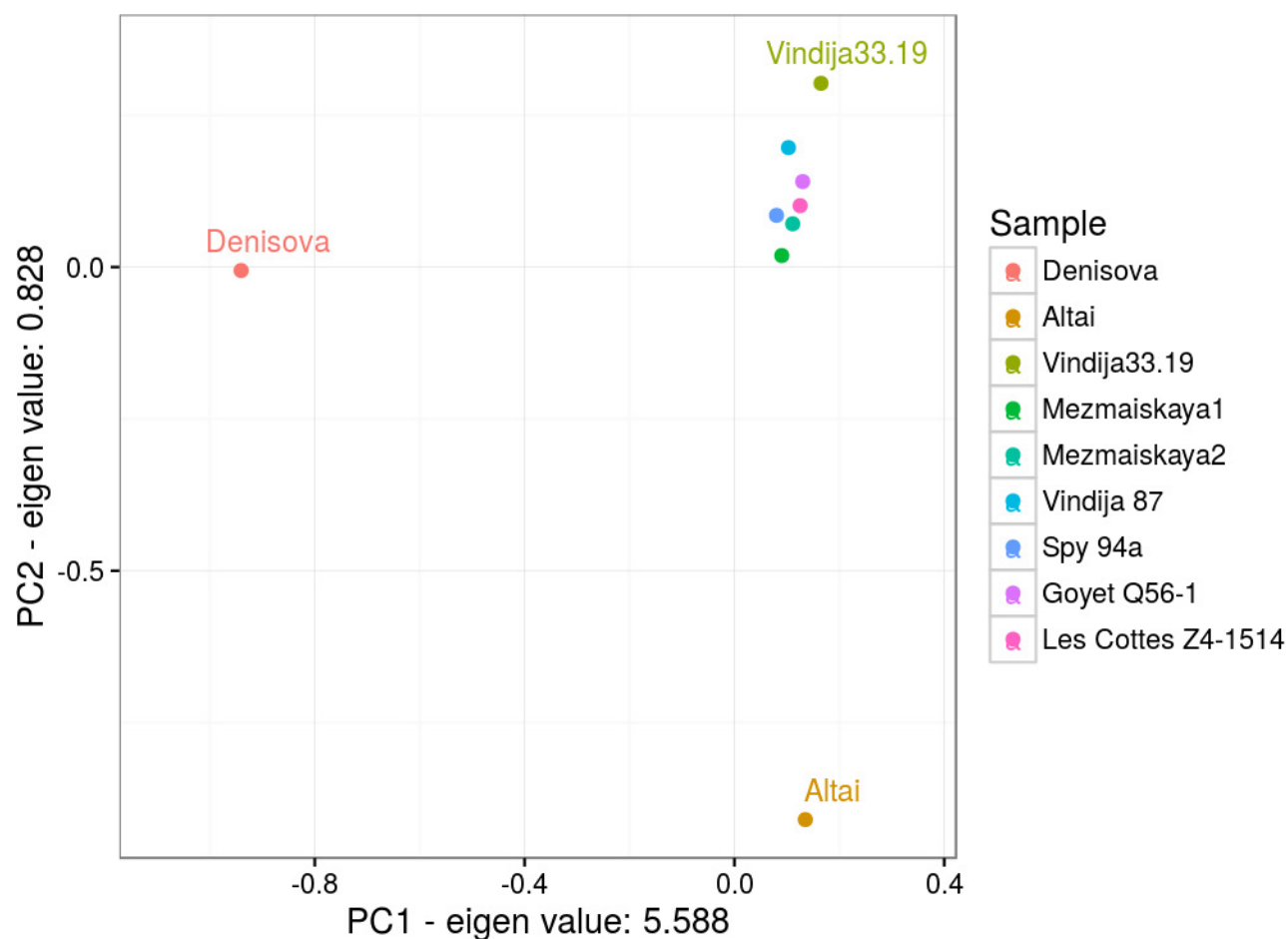
Extended Data Figure 2 | Fragment size distribution of fragments longer than 35 bp mapped to the human reference genome with $MQ \geq 25$ for each of the five late Neanderthals. All fragments are

depicted in solid lines and fragments with C-to-T substitutions to the reference genome (putatively deaminated fragments) are depicted with dashed lines.



Extended Data Figure 3 | Sex determination based on the number of fragments aligning to the X chromosome and the autosomes. The expected ratios of X to (X + autosomal) fragments for a female and a male

individual are depicted as dashed lines. The results were concordant for all fragments (in red) and for deaminated fragments only (in grey).



Extended Data Figure 4 | Principal component analysis of the genomes of Vindija 33.19, Altai, the Denisovan individual, five late Neanderthals and Mezmaiskaya 1. Genomes of the high-coverage archaics were used to estimate the eigenvectors of the genetic variation and

low-coverage Neanderthals were projected onto the plane. Only transversion polymorphisms and bi-allelic sites were considered for the analysis, to a total of 1,010,417 sites as defined by the high-coverage genomes. PC, principal component.

Extended Data Table 1 | Amount of data generated for Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87 and Spy 94a

Specimen	All fragments					Fragments with terminal C-to-T substitutions to the reference genome		
	Number of sequenced fragments	Number of fragments ≥ 35 bp	Number of mapped fragments ≥ 35 bp, MQ ≥ 25 , Map35_100%	Number of unique fragments ≥ 35 bp, MQ ≥ 25 , Map35_100%	Obtained nuclear coverage	Number of mapped fragments ≥ 35 bp, MQ ≥ 25 , Map35_100%	Number of unique fragments ≥ 35 bp, MQ ≥ 25 , Map35_100%	Obtained nuclear coverage
Les Cottés Z4-1514	1,776,108,129	978,981,412	253,134,472	121,336,498	2.71	59,777,476	41,804,332	1.00
Goyet Q56-1	917,798,995	575,597,226	124,582,445	80,515,019	2.18	13,585,646	9,859,280	0.27
Mezmaiskaya 2	2,571,128,011	1,356,302,134	125,815,832	74,407,074	1.74	32,838,281	23,521,855	0.56
Vindija 87	1,353,546,508	678,005,556	110,385,510	60,034,976	1.27	32,597,516	22,828,056	0.49
Spy 94a	733,212,063	278,595,622	72,962,863	46,836,553	1.00	14,301,192	10,003,560	0.22

The number of fragments after merging all of the sequencing libraries together is shown. The coverage of the nuclear genomes is determined by counting the number of bases with a base quality of at least 30 in fragments longer than 35 bp with MQ ≥ 25 that overlapped highly mappable regions of autosomes of the human genome, and dividing that number by the total length of these regions.

Extended Data Table 2 | Relationship of the late Neanderthals and Mezmaiskaya 1 to the Altai and Vindija 33.19 Neanderthals calculated as $D(\text{Altai, Vindija 33.19, Neanderthal, outgroup})$ for all fragments and deaminated fragments, restricted to transversions

D (Altai, Vindija 33.19, Neanderthal, outgroup)				All fragments				Deaminated fragments			
W	X	Y	Z	% D	Z-score	BABA	ABBA	% D	Z-score	BABA	ABBA
Altai	Vindija33.19	Les Cottés Z4-1514	Dinka	48.87	-55.31	30,276	10,399	49.08	-52.83	19,048	6,506
Altai	Vindija33.19	Les Cottés Z4-1514	Mbuti	49.08	-57.83	30,340	10,362	49.34	-55.25	19,077	6,471
Altai	Vindija33.19	Les Cottés Z4-1514	Chimp	49.16	-59.83	29,427	10,028	49.38	-56.69	18,597	6,302
Altai	Vindija33.19	Les Cottés Z4-1514	Orang	48.88	-60.25	28,149	9,664	49.15	-57.21	17,775	6,060
Altai	Vindija33.19	Les Cottés Z4-1514	Rhesus	48.35	-58.96	25,856	9,003	48.38	-55.77	16,281	5,664
Altai	Vindija33.19	Goyet Q56-1	Dinka	55.87	-66.92	33,208	9,403	56.32	-56.88	8,673	2,423
Altai	Vindija33.19	Goyet Q56-1	Mbuti	56.00	-68.27	33,269	9,383	56.32	-58.37	8,689	2,428
Altai	Vindija33.19	Goyet Q56-1	Chimp	55.96	-68.97	32,188	9,090	56.76	-59.25	8,485	2,340
Altai	Vindija33.19	Goyet Q56-1	Orang	55.68	-68.11	30,735	8,749	56.79	-59.39	8,125	2,239
Altai	Vindija33.19	Goyet Q56-1	Rhesus	55.40	-67.04	28,312	8,125	56.31	-58.09	7,483	2,092
Altai	Vindija33.19	Mezmaiskaya 1	Dinka	36.58	-36.63	20,392	9,468	37.09	-32.69	8,379	3,845
Altai	Vindija33.19	Mezmaiskaya 1	Mbuti	36.80	-37.33	20,516	9,479	37.43	-33.41	8,438	3,842
Altai	Vindija33.19	Mezmaiskaya 1	Chimp	36.70	-38.66	19,860	9,197	37.00	-33.45	8,126	3,736
Altai	Vindija33.19	Mezmaiskaya 1	Orang	36.18	-37.38	18,904	8,860	36.77	-32.27	7,758	3,586
Altai	Vindija33.19	Mezmaiskaya 1	Rhesus	36.00	-37.77	17,433	8,204	36.36	-32.05	7,129	3,327
Altai	Vindija33.19	Mezmaiskaya 2	Dinka	47.44	-42.97	26,802	9,553	46.84	-40.59	12,477	4,516
Altai	Vindija33.19	Mezmaiskaya 2	Mbuti	47.83	-47.39	26,781	9,451	47.13	-44.43	12,471	4,481
Altai	Vindija33.19	Mezmaiskaya 2	Chimp	47.92	-51.11	26,019	9,160	47.37	-47.06	12,233	4,369
Altai	Vindija33.19	Mezmaiskaya 2	Orang	47.86	-53.48	24,921	8,788	47.07	-47.53	11,700	4,210
Altai	Vindija33.19	Mezmaiskaya 2	Rhesus	47.33	-49.32	22,908	8,189	46.65	-44.74	10,799	3,929
Altai	Vindija33.19	Spy 94a	Dinka	54.86	-56.62	21,416	6,242	55.17	-50.39	6,558	1,895
Altai	Vindija33.19	Spy 94a	Mbuti	55.18	-59.76	21,453	6,196	55.53	-52.89	6,567	1,878
Altai	Vindija33.19	Spy 94a	Chimp	55.32	-60.58	20,814	5,988	55.39	-50.67	6,354	1,824
Altai	Vindija33.19	Spy 94a	Orang	55.01	-60.26	19,793	5,746	55.33	-50.10	6,052	1,740
Altai	Vindija33.19	Spy 94a	Rhesus	54.46	-58.94	18,179	5,360	54.93	-48.52	5,584	1,624

All late Neanderthals and Mezmaiskaya 1 are significantly closer to Vindija 33.19 than to the Altai Neanderthal, irrespective of the outgroup used. Blue denotes $Z \ll -2$. A total of 1,567,449 sites that are transversions among Neanderthals and for which at least one late Neanderthal has coverage were used. Chimp, chimpanzee; Orang, orangutan; Rhesus, rhesus macaque.

Extended Data Table 3 | The fraction of derived alleles among putatively deaminated fragments that each of the low-coverage individuals shares with the Altai Neanderthal, Vindija 33.19, the Denisovan individual and a present-day human genome

	Human (%)	Neanderthal (%)	Altai Neanderthal (%)	Vindija 33.19 (%)	Denisovan (%)	Neanderthal-Denisova (%)	Human-Neanderthal (%)	Human-Denisova (%)
Les Cottés Z4-1514	0.67 [0.65-0.70]	93.01 [92.89-93.13]	18.00 [17.69-18.32]	46.41 [46.06-46.77]	0.86 [0.82-0.89]	97.90 [97.83-97.98]	97.30 [97.17-97.43]	3.22 [3.07-3.38]
Goyet Q56-1	0.60 [0.56-0.65]	94.38 [94.19-94.57]	16.47 [15.93-17.02]	53.55 [52.91-54.18]	0.80 [0.74-0.86]	98.17 [98.04-98.29]	97.72 [97.49-97.92]	2.52 [2.29-2.78]
Spy 94a	0.71 [0.66-0.77]	93.89 [93.65-94.12]	16.52 [15.90-17.17]	51.45 [50.70-52.21]	0.80 [0.73-0.87]	97.83 [97.66-97.99]	97.62 [97.35-97.87]	3.06 [2.75-3.39]
Vindija 87	0.36 [0.33-0.38]	97.08 [96.97-97.19]	6.81 [6.53-7.09]	75.93 [75.49-76.36]	0.37 [0.34-0.40]	99.05 [98.98-99.12]	98.91 [98.79-99.02]	1.38 [1.24-1.53]
Mezmaiskaya 2	0.65 [0.62-0.69]	92.78 [92.62-92.93]	17.91 [17.51-18.31]	44.65 [44.19-45.12]	0.83 [0.79-0.88]	97.74 [97.63-97.84]	97.53 [97.36-97.68]	3.38 [3.18-3.59]
Mezmaiskaya 1	0.76 [0.72-0.80]	91.87 [91.67-92.06]	20.30 [19.81-20.80]	38.38 [37.85-38.91]	0.97 [0.91-1.03]	97.61 [97.48-97.73]	96.80 [96.58-97.00]	3.73 [3.49-4.00]
Vindija 33.16	3.60 [3.50-3.71]	93.39 [93.18-93.59]	17.09 [16.55-17.65]	58.99 [58.36-59.63]	3.94 [3.81-4.08]	96.14 [95.95-96.32]	95.46 [95.15-95.74]	4.89 [4.56-5.24]
Vindija 33.25	3.04 [2.94-3.15]	94.17 [93.95-94.37]	16.69 [16.10-17.30]	60.81 [60.11-61.49]	3.41 [3.28-3.55]	96.53 [96.33-96.72]	95.79 [95.46-96.09]	3.77 [3.46-4.12]
Vindija 33.26	3.49 [3.38-3.62]	93.54 [93.31-93.77]	17.04 [16.43-17.67]	59.65 [58.94-60.36]	3.70 [3.55-3.85]	96.25 [96.04-96.45]	95.40 [95.04-95.72]	4.56 [4.21-4.95]
Feldhofer 1	4.71 [2.40-9.01]	94.38 [87.51-97.58]	20 [8.86-39.13]	50.00 [34.07-65.93]	1.80 [0.50-6.33]	90.63 [81.02-95.63]	96.00 [80.46-99.29]	0 [0-12.87]
El Sidron 1253	1.29 [0.44-3.72]	90.99 [84.21-95.03]	17.50 [8.75-31.95]	42.86 [30.02-56.73]	4.35 [2.22-8.34]	95.24 [88.39-98.13]	97.44 [86.82-99.55]	2.33 [0.41-12.06]
Denisova 4	2.63 [1.51-4.54]	2.31 [1.06-4.94]	2.13 [0.59-7.43]	2.04 [0.56-7.14]	71.43 [65.87-76.40]	96.15 [91.86-98.23]	10.00 [5.15-18.51]	90.32 [80.45-95.49]
Denisova 8	1.60 [1.39-1.85]	6.75 [6.13-7.43]	1.75 [1.23-2.48]	1.83 [1.37-2.45]	60.02 [58.92-61.12]	92.53 [91.66-93.32]	15.32 [13.72-17.06]	87.72 [86.01-89.25]

The state of DNA fragments overlapping the positions at which the high-coverage genomes of the Altai Neanderthal, the Vindija 33.19 Neanderthal, the Denisovan individual and a present-day African (Mbuti, HGDP00982) differ from those of the great apes were investigated. Fragments longer than 35 bp with MQ ≥ 25 and within the highly mappable regions of the genome that had terminal C-to-T substitutions reported in the Supplementary Table 3.2 were used. 95% binomial confidence intervals are provided in brackets.

Extended Data Table 4 | Time of separation of late Neanderthals and Mezmaiskaya 1 (A) from the ancestor with the high-coverage genomes of Altai and Vindija 33.19 Neanderthals, Denisovan individual and a present-day human (B), when measured in terms of time of split from the B individual (split A–B), or time from present (split-time + branch shortening)

A	B	All fragments				Deaminated fragments			
		% F(A B)	Split A-B (ky)	Split time + bs (kya)	95% CI (kya)	% F(A B)	Split A-B (ky)	Split time + bs (kya)	95% CI (kya)
Les Cottés Z4-1514	Altai	35.2	35.0	157.5	144.6–160.7	34.7	37.8	160.2	156.4–163.5
Goyet Q56-1	Altai	35.8	22.3	144.7	142.8–157.6	35.9	22.2	144.6	141.5–159.3
Mezmaiskaya 2	Altai	35.4	33.8	156.2	143.9–159.7	34.2	40.2	162.6	158.7–166.5
Spy 94a	Altai	34.4	39.2	161.7	158.2–165.2	33.6	43.1	165.5	160.4–186.2
Vindija 87	Altai	35.2	35.4	157.8	144.7–161.1	34.6	38.3	160.8	156.2–164.8
Mezmaiskaya 1	Altai	34.8	37.4	159.9	156.1–163.0	34.4	42.9	165.3	161.1–178.6
Les Cottés Z4-1514	Vindija 33.19	35.3	15.1	66.9	66.1–68.7	34.3	18.3	70.0	67.9–72.4
Goyet Q56-1	Vindija 33.19	37.8	8.3	60.1	58.8–61.7	38.0	8.1	59.9	57.8–61.9
Mezmaiskaya 2	Vindija 33.19	35.6	14.3	66.1	64.3–67.7	34.3	18.3	70.0	67.4–74.5
Spy 94a	Vindija 33.19	36.9	10.4	62.2	60.6–64.3	37.2	9.9	62.7	59.1–64.8
Vindija 87	Vindija 33.19	46.8	0.0	51.8	51.8–51.8	45.0	0.0	61.7	51.8–51.8
Mezmaiskaya 1	Vindija 33.19	31.6	42.0	93.8	81.5–96.9	30.6	46.9	98.7	94.5–102.9
Les Cottés Z4-1514	Denisovan	12.6	333.1	405.1	381.9–448.2	12.2	346.7	418.7	390.4–470.8
Goyet Q56-1	Denisovan	12.8	328.6	400.6	377.3–438.0	13.1	322.4	394.3	367.1–434.0
Mezmaiskaya 2	Denisovan	12.4	338.2	410.2	387.0–458.4	11.9	362.0	434.0	398.9–483.8
Spy 94a	Denisovan	12.5	336.0	407.9	384.7–451.6	12.4	339.4	411.3	382.4–470.8
Vindija 87	Denisovan	12.5	337.1	409.1	385.8–455.0	12.2	346.7	418.7	390.9–470.2
Mezmaiskaya 1	Denisovan	12.3	341.6	413.6	389.2–463.5	12.3	345.0	417.0	388.7–471.4
Les Cottés Z4-1514	Mbuti	17.5	525.9	525.9	511.3–539.8	17.2	549.3	549.3	533.2–565.4
Goyet Q56-1	Mbuti	17.7	515.0	515.0	501.8–528.9	17.6	520.8	520.8	502.6–540.6
Mezmaiskaya 2	Mbuti	17.6	521.6	521.6	506.9–536.2	17.4	535.4	535.4	517.9–553.0
Spy 94a	Mbuti	18.0	502.6	502.6	488.7–517.2	17.3	542.7	542.7	518.6–566.1
Vindija 87	Mbuti	17.6	523.0	523.0	509.1–537.6	17.3	542.0	542.0	524.5–558.8
Mezmaiskaya 1	Mbuti	17.8	512.8	512.8	498.9–528.1	17.1	552.2	552.2	534.7–569.8

Results reported for all fragments and for deaminated fragments only (Supplementary Tables 3.1, 3.2). 95% confidence intervals for $F(A|B)$ values are estimated via weighted block jackknife, using blocks of 5 Mb. In order to obtain estimates of split times from present, we correct 'split A–B' by adding the age of the high-coverage B-individual estimated from branch-shortening (column: split time + branch shortening (ka)).

Moving magnetoencephalography towards real-world applications with a wearable system

Elena Boto^{1*}, Niall Holmes^{1*}, James Leggett^{1*}, Gillian Roberts^{1*}, Vishal Shah², Sofie S. Meyer^{3,4}, Leonardo Duque Muñoz³, Karen J. Mullinger^{1,5}, Tim M. Tierney³, Sven Bestmann^{3,6}, Gareth R. Barnes^{3§}, Richard Bowtell^{1§} & Matthew J. Brookes^{1§}

Imaging human brain function with techniques such as magnetoencephalography¹ typically requires a subject to perform tasks while their head remains still within a restrictive scanner. This artificial environment makes the technique inaccessible to many people, and limits the experimental questions that can be addressed. For example, it has been difficult to apply neuroimaging to investigation of the neural substrates of cognitive development in babies and children, or to study processes in adults that require unconstrained head movement (such as spatial navigation). Here we describe a magnetoencephalography system that can be worn like a helmet, allowing free and natural movement during scanning. This is possible owing to the integration of quantum sensors^{2,3}, which do not rely on superconducting technology, with a system for nulling background magnetic fields. We demonstrate human electrophysiological measurement at millisecond resolution while subjects make natural movements, including head nodding,

stretching, drinking and playing a ball game. Our results compare well to those of the current state-of-the-art, even when subjects make large head movements. The system opens up new possibilities for scanning any subject or patient group, with myriad applications such as characterization of the neurodevelopmental connectome, imaging subjects moving naturally in a virtual environment and investigating the pathophysiology of movement disorders.

Magnetoencephalography¹ (MEG) allows direct imaging of human brain electrophysiology by measurement of magnetic fields generated at the scalp by neural currents. Mathematical analysis of those fields enables the generation of 3D images that show the formation and dissolution of brain networks in real time. MEG measurements of brain activity are currently made using an array of superconducting sensors placed around the head^{1,4}. These cryogenically cooled sensors have femtotesla (fT) sensitivity, which is needed to detect the weak magnetic fields produced by the brain. Unfortunately, the requirement

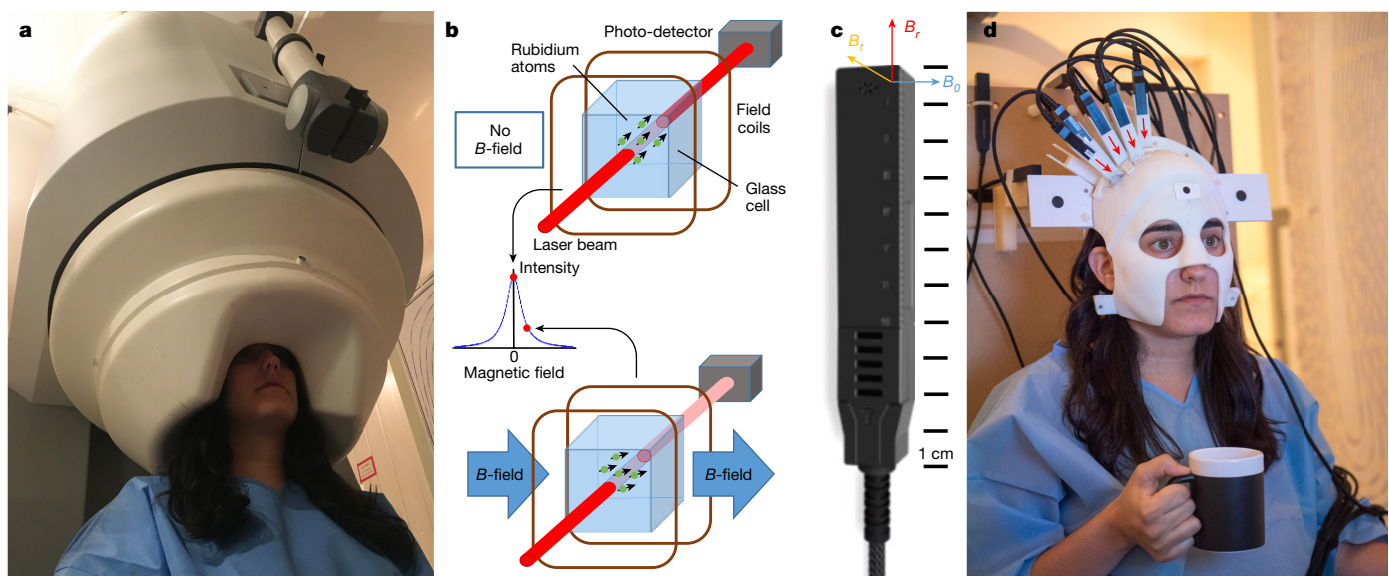


Figure 1 | A new generation of MEG system. **a**, A conventional 275-channel cryogenic MEG system. Weighing about 450 kg, the system is fixed and cumbersome and subjects must remain still relative to the fixed sensor array. **b**, Schematic illustration of zero-field resonance in an OPM sensor. Top, operation in zero-field; bottom, Larmor precession when an external field (B -field) impinges on the cell and the transmitted light intensity is reduced. **c**, A commercial OPM sensor made by QuSpin. The geometry used is illustrated by the coloured axes where B_r is the radial

field component, B_t the tangential field component and B_o the direction along which the laser beam is oriented. **d**, Our prototype OPM-MEG system helmet. The helmet weighs 905 g and is customized so that the sensors (which in this prototype cover only the right sensorimotor cortex) are directly adjacent to the scalp surface. The subject is free to move their head. The measured radial field direction for the sensors is illustrated by the red arrows.

¹Sir Peter Mansfield Imaging Centre, School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK. ²QuSpin Inc., 331 South 104th Street, Suite 130, Louisville, Colorado 80027, USA. ³Wellcome Centre for Human Neuroimaging, UCL Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, UK. ⁴Institute of Cognitive Neuroscience, University College London, 17–19 Queen Square, London WC1N 3AZ, UK. ⁵Centre for Human Brain Health, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ⁶Sobell Department for Motor Neuroscience and Movement Disorders, UCL Institute of Neurology, University College London, Queen Square House, Queen Square, London WC1N 3BG, UK.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

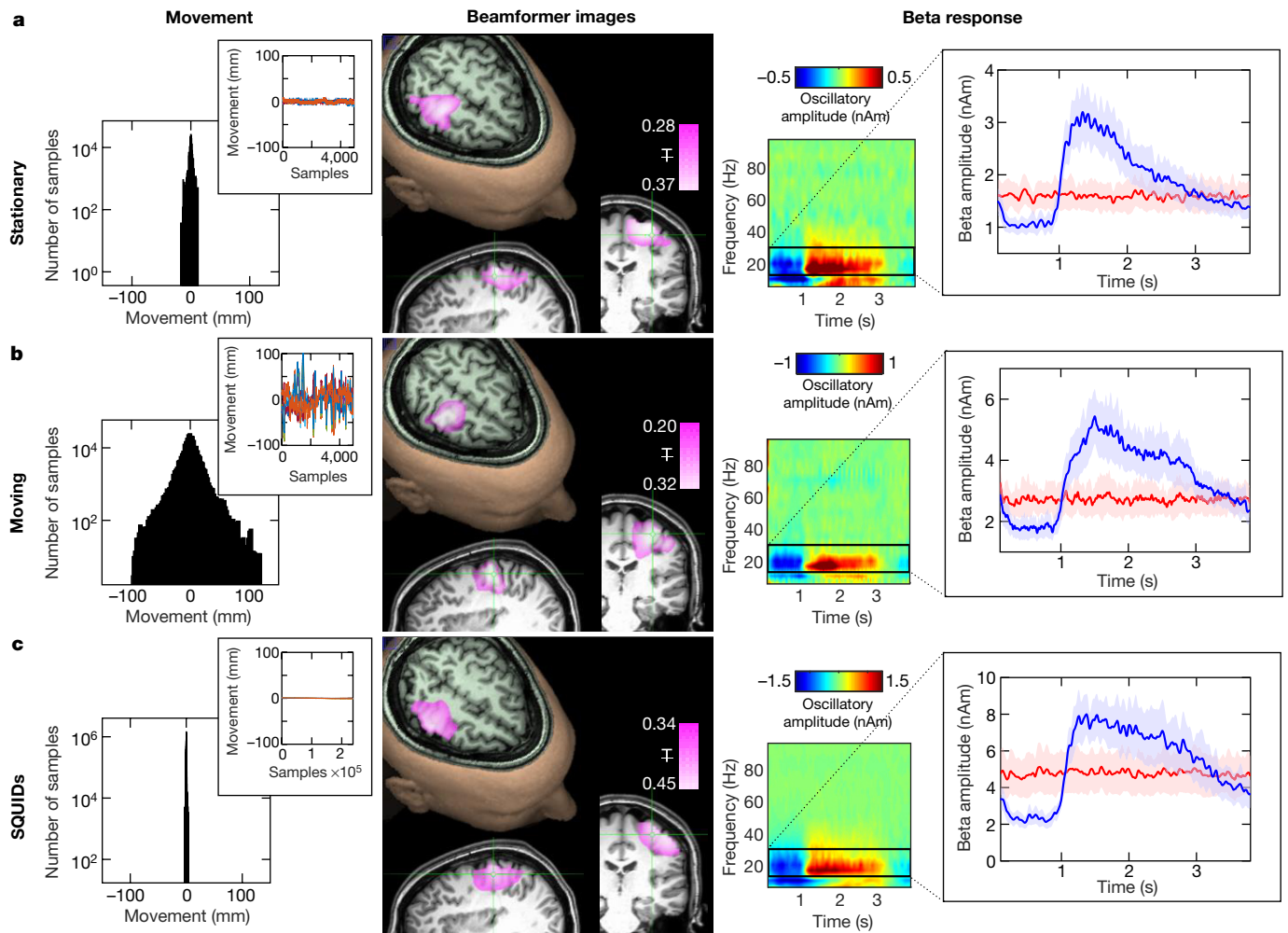


Figure 2 | OPM-MEG results. **a**, Results when the subject was asked to remain still. **b**, Results when the subject was moving. **c**, Data collected using a cryogenic MEG instrument (SQUIDS) for comparison. In each row, the left panel contains a histogram showing movement of three fiducial markers on the subject's head. The inset graph shows the change in marker positions over a typical experiment; different colours show movement in three Cartesian axes of the three markers. The middle panel shows the change in beta band power due to finger abduction (pink) overlaid onto axial, sagittal and coronal slices of the anatomical MRI. The functionally active region overlays contralateral sensorimotor cortex. In the right panel, a time–frequency spectrogram depicts changes in

neural oscillations during finger abduction. The inset graph shows the characteristic beta band response for finger abduction (blue) and rest trials (red). In all cases the results are averaged over trials and experiments and the shaded region shows s.e. over six experiments. SIR ranged from 4.3 to 8.2 for static OPM measures, 4.2 to 5.8 for moving OPMs and 4.9 to 7.9 for the cryogenic system. Further analysis (Extended Data Fig. 2 and Supplementary Information section 2) showed that the OPM system outperformed the cryogenic system in terms of both spatial resolution and robustness across experiments. Temporal resolution was quantified at 130 Hz.

for cooling means that sensors must be housed within a liquid helium dewar with a vacuum space separating sensors from the scalp. MEG systems are therefore cumbersome (Fig. 1a) and sensor positions are fixed in a one-size-fits-all helmet. Any motion of the head relative to the sensors reduces data quality markedly: even a 5-mm movement can prevent use of the data⁵. Further, the brain-to-sensor distance, which is substantial in adults (about 3 cm), is increased markedly in subjects with small heads, reducing the available signal because the magnetic field decreases with the square of the distance between the source and the sensor. These characteristics make participation in MEG studies challenging for many subject groups, including infants and many patients. They also make the MEG scanner environment unnatural, and limit the experimental paradigms that can be studied. Here, we describe a transformative MEG technology that can be worn on the head during movement. This opens up the possibility for non-invasive mapping of human electrophysiology across all ages and patient groups, with subjects who are free to move and interact with the real world.

At the core of our system is an array of optically pumped magnetometers (OPMs)—magnetic field sensors that rely on the atomic

properties of alkali metals. These sensors have undergone marked development in recent years^{2,3,6–9} and are well suited to MEG measurements^{10–16}. In our system, each sensor contains a $3 \times 3 \times 3\text{-mm}^3$ glass cell containing ^{87}Rb vapour, which is heated to approximately 150°C . A 795-nm circularly polarized laser beam, tuned to the D1 transition of rubidium, is used to spin-polarize the atoms, and the intensity of laser light transmitted through the cell is detected using a photodiode. In zero magnetic field, the spin magnetic moments align with the beam, and transmission of laser light to the photodiode is maximized. However, a magnetic field perpendicular to the beam causes Larmor precession, rotating the magnetic moments away from alignment. This causes a measurable drop in light transmission. The resulting effect is a zero-field resonance (Fig. 1b), which acts as a sensitive magnetic field indicator.

Each sensor is an integrated unit (Fig. 1c) with a noise level comparable to that of a superconducting quantum interference device (SQUID; about $15\text{ fT per } \sqrt{\text{Hz}}$) and a dynamic range of $\pm 1.5\text{ nT}$. Although the cell is heated, sensors can be mounted on the scalp because their external surfaces remain close to body temperature. Our prototype system (Fig. 1d)

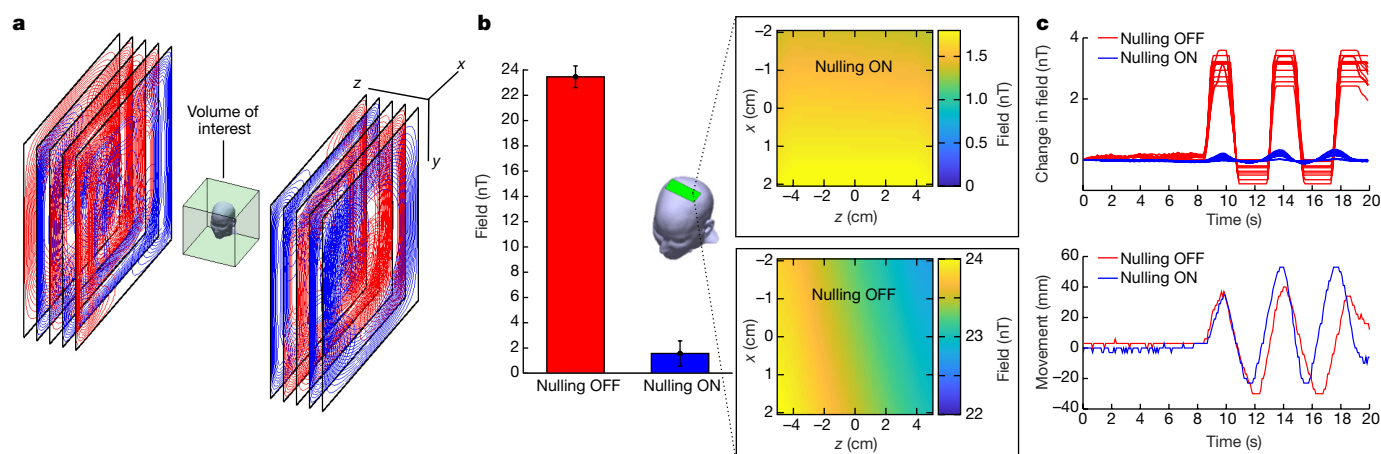


Figure 3 | Biplanar fingerprint-coil system for removing remnant static magnetic fields. **a**, Schematic of the coils, which are confined to two planes surrounding a $40 \times 40 \times 40\text{-cm}^3$ region of interest in which the head is allowed to move. The five separate layers represent wire paths that generate fields B_x , B_y , B_z , dB_x/dz and dB_z/dz . **b**, Bar chart showing field magnitude with and without field nulling. Inset images show spatial field variation of B_x . The static field was reduced from $23 \pm 1\text{ nT}$ to $1.6 \pm 1.0\text{ nT}$. dB_x/dz measured across a 10-cm region spanning the width

of the head (green box in **b**) was reduced from 10 nT m^{-1} to 0.28 nT m^{-1} . **c**, Top, output of 13 OPMs over time while the subject nods their head (bottom, head movement). Blue and red traces show the cases where the field nulling system was on and off, respectively. Note that without field nulling the OPMs saturate during head movement, whereas with nulling the sensors continue to work, leaving an artefact that is comparable in magnitude to that due to an eye-blink.

comprised an array of sensors that were mounted in a 3D-printed 'scanner-cast'. The scanner-cast¹² was designed using an anatomical magnetic resonance imaging (MRI) scan, such that the internal surface snugly fits the subject's head, while the external surface accommodates the OPMs, which were positioned over the right sensorimotor cortex. Four additional reference sensors, sited near the subject's head, were used to measure background interference.

As a first demonstration, we measured electrophysiological activity in the right sensorimotor cortex during visually cued abduction of the left index finger. This task robustly elicits a reduction in endogenous beta band (13–30 Hz) oscillations during movement and a rebound (increase above baseline) following movement cessation¹⁷. Although simple, this task has been applied widely, with beta modulation being used as a marker of brain plasticity^{18,19}, psychosis^{20,21} and white matter degradation²². A single experiment comprised 50 trials, each involving 1 s of finger abduction and 3 s of rest. Fifty 'dummy' trials (during which the subject did nothing) were interleaved to allow an estimate of baseline activity. A subject undertook the experiment twelve times: six during which they kept as still as possible, and six during which they made natural head movements, including nodding and shaking, stretching and drinking. Head motion was measured using an Xbox Kinect camera, which tracked the movement of three fiducial markers on the head. We also undertook the same static set of experiments using a cryogenic MEG system.

Figure 2 shows OPM-MEG data measured when the subject kept still (Fig. 2a) and moved (Fig. 2b). Subject motion is shown in the left-hand column. In the static case, motion was less than $\pm 1\text{ cm}$, whereas in the moving case it exceeded $\pm 10\text{ cm}$. The middle column shows images^{23,24} of beta modulation (pink overlay) and the right-hand column shows a time–frequency spectrogram (TFS) of oscillatory change. In the TFS, blue indicates a decrease in oscillatory amplitude relative to baseline whereas red indicates an increase. Line plots of beta amplitude are shown in the inset. Equivalent data from the cryogenic system are shown in Fig. 2c, where movement was (necessarily) constrained to less than 2 mm. OPM-MEG performed consistently across experiments with the characteristic beta decrease and post movement rebound clearly delineated and localized to the sensorimotor cortex. Despite an order of magnitude increase in head movement, there was no significant difference in signal-to-interference ratio (SIR) between the moving and static runs ($P = 0.39$; two-sided Wilcoxon sum-rank test) and no correlation between the degree of movement and response

size (Extended Data Fig. 1 and Supplementary Information section 1). The spatial resolution of the OPM system was better than that of the cryogenic system, even with only 13 sensors (Extended Data Fig. 2 and Supplementary Information section 2). These data, along with a similar analysis of evoked responses (Extended Data Fig. 3 and Supplementary Information section 3), show clearly that the wearable system can be used to collect high-fidelity data even in the presence of large head movements.

Critical to the wearable MEG system is a means to null the background static magnetic field impinging on the OPMs. The system is housed inside a magnetically shielded room (MSR). However, the remnant Earth's field in the MSR is about 25 nT, and it is spatially inhomogeneous. Any sensor movement through this field during an MEG recording would result in a field change much larger than the fields of interest, and would exceed the narrow ($\pm 1.5\text{ nT}$) operational range of the OPMs, rendering them inoperable. In addition, such changes can modulate the sensor gain (Extended Data Fig. 4 and Supplementary Information section 4). To ameliorate this problem, we constructed a set of bi-planar electromagnetic coils designed to generate fields equal and opposite to the remnant Earth's field, thereby cancelling it out.

The coils were designed^{25,26} on two $1.6 \times 1.6\text{-m}^2$ planes, placed either side of the subject with a 1.5-m separation (Fig. 3a). Three coils generated spatially uniform fields (B_x , B_y and B_z) while two additional coils were used to remove the dominant field variations (dB_x/dz and dB_z/dz). In this way, unlike standard field-nulling technologies (for example, tri-axial Helmholtz coils), our system can account for spatial variation of the field over a $40 \times 40 \times 40\text{-cm}^3$ volume of interest enclosing the head. Furthermore, we were able to cancel all components of the field vector using coils confined to just two planes, hence retaining easy access to the subject. Four reference OPM sensors were coupled to the coils in a feedback loop to null the residual static field in the volume of interest. We achieved a 15-fold reduction in the remnant Earth's field and a 35-fold reduction in the dominant field gradient (Fig. 3b). Figure 3c shows OPM measurements made during 7-cm head movements, with (blue) and without (red) field nulling. The results show that without field nulling, the OPM sensors failed (evidenced by the saturation). With field nulling, however, the OPMs were able to capture MEG data even while the head was moving (Supplementary Information section 5).

The ability to map human electrophysiology non-invasively, with whole brain coverage and high spatiotemporal resolution is, to our

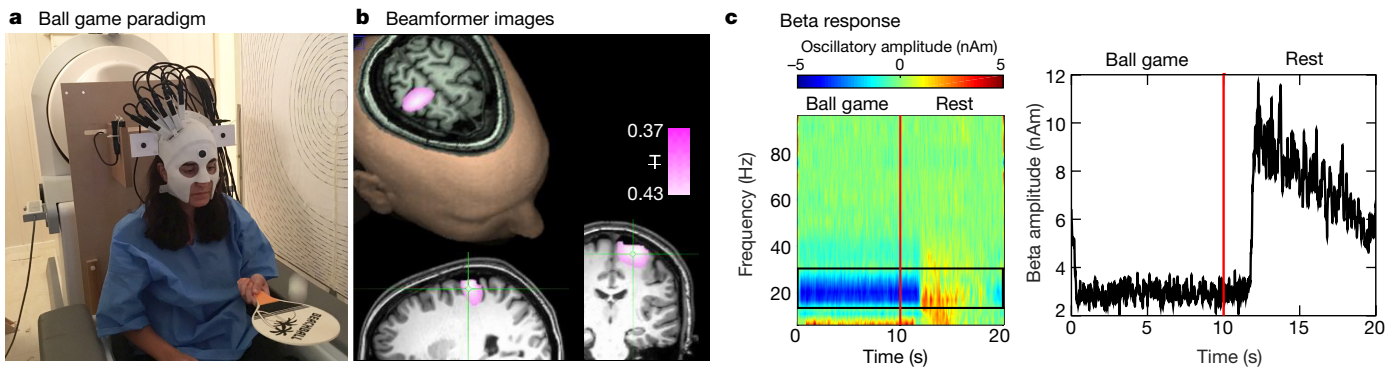


Figure 4 | An example ‘real world’ imaging paradigm. **a**, Experimental set-up: the seated subject continually bounced a table tennis ball off a bat for 10 s. This was followed by a 10-s baseline period, during which they did nothing. This was repeated 29 times. **b**, Spatial signature of beta band oscillatory change during periods of playing the ball game compared to rest. Note the difference in localization compared to Fig. 2, with the beta

knowledge, unique to MEG²⁷. Alternative techniques either lack spatial specificity (for example, electroencephalography (EEG), which is also more susceptible to muscle artefacts (see Extended Data Fig. 5 and Supplementary Information Section 6)) or provide only indirect access to brain function via metabolic processes (for example, functional magnetic resonance imaging (fMRI)). However, current MEG systems exclude many subject cohorts and experimental paradigms. Our system represents a step change for functional imaging. A wearable instrument with scalp-mounted sensors means that subjects can be scanned at all ages, from babies to elderly patients, allowing imaging during key stages of cognitive development and decline. For example, it is well known that efficient communication between spatially separate cortical regions is key to healthy brain function, and that neural oscillations help to mediate such connectivity^{28,29}. However, little is known about how electrophysiological brain networks emerge during the early years of life. Our system can characterize those networks, and the spectro-temporal profile of the connectivities that underlie them (see Extended Data Fig. 6 and Supplementary Information section 7). This highlights the potential utility of OPM-MEG for characterizing the developmental connectome.

A wearable system also opens doors to myriad neuroscientific investigations in which subjects can move naturally and interact with the real world. An example is given in Fig. 4, which shows brain activity elicited when the subject played a ball game (Fig. 4a). This naturalistic paradigm requires free, rapid and unpredictable head and arm movement; nevertheless, we were able to localize changes in beta oscillations to the arm and wrist area of motor cortex (Fig. 4b) and track the dynamics of oscillatory modulation (Fig. 4c). To evaluate the robustness of this measure, similar data were captured in two further subjects. Task-induced beta modulation, relative to baseline, was measured at 61%, 80% and 70% in the three participants. This simple example demonstrates the power of wearable neuroimaging to accurately assess brain function during a real world visuomotor coordination task, adding a new dimension to previous work³⁰. This example could readily be extended to examine, for example, the neural correlates of motor coordination, their maturation during neurodevelopment and their breakdown in movement disorders. Such studies using naturalistic stimuli are inaccessible to conventional neuroimaging scanners owing to the movement required to undertake the task. In sum, this technology has transformative potential across a range of neuroscientific and clinical applications where knowledge of brain electrophysiology is informative.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August 2017; accepted 13 February 2018.

Published online 21 March 2018.

modulation localized to the arm and wrist area of sensorimotor cortex (distinct from the hand area in Fig. 2). **c**, Trial averaged time-frequency spectrogram (left) and amplitude of beta oscillations averaged over trials (right). The maximum head movement during this paradigm, assessed by the Kinect camera, was 6 cm.

- Cohen, D. Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer. *Science* **175**, 664–666 (1972).
- Kominis, I. K., Kornack, T. W., Allred, J. C. & Romalis, M. V. A subfemtotesla multichannel atomic magnetometer. *Nature* **422**, 596–599 (2003).
- Shah, V. K. & Wakai, R. T. A compact, high performance atomic magnetometer for biomedical applications. *Phys. Med. Biol.* **58**, 8153–8161 (2013).
- Hämäläinen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasma, O. V. Magnetoencephalography: theory, instrumentation, and applications to non-invasive studies of the working human brain. *Rev. Mod. Phys.* **65**, 413–497 (1993).
- Gross, J. *et al.* Good practice for conducting and reporting MEG research. *Neuroimage* **65**, 349–363 (2013).
- Allred, J. C., Lyman, R. N., Kornack, T. W. & Romalis, M. V. High-sensitivity atomic magnetometer unaffected by spin-exchange relaxation. *Phys. Rev. Lett.* **89**, 130801 (2002).
- Shah, V., Knappe, S., Schwindt, P. D. D. & Kitching, J. Subpicotesla atomic magnetometry with a microfabricated vapour cell. *Nat. Photon.* **1**, 649–652 (2007).
- Dupont-Roc, J., Haroche, S. & Cohen-Tannoudji, C. Detection of very weak magnetic fields (10–9 gauss) by ⁸⁷Rb zero-field level crossing resonances. *Phys. Lett. A* **28**, 638–639 (1969).
- Borna, A. *et al.* A 20-channel magnetoencephalography system based on optically pumped magnetometers. *Phys. Med. Biol.* **62**, 8909–8923 (2017).
- Alem, O., Benison, A. M., Barth, D. S., Kitching, J. & Knappe, S. Magnetoencephalography of epilepsy with a microfabricated atomic magnetometer. *J. Neurosci.* **34**, 14324–14327 (2014).
- Alem, O. *et al.* Magnetic field imaging with microfabricated optically-pumped magnetometers. *Opt. Express* **25**, 7849–7858 (2017).
- Boto, E. *et al.* A new generation of magnetoencephalography: room temperature measurements using optically-pumped magnetometers. *Neuroimage* **149**, 404–414 (2017).
- Johnson, C., Schwindt, P. D. D. & Weisend, M. Magnetoencephalography with a two-color pump-probe, fiber-coupled atomic magnetometer. *Appl. Phys. Lett.* **97**, 243703 (2010).
- Johnson, C. N., Schwindt, P. D. D. & Weisend, M. Multi-sensor magnetoencephalography with atomic magnetometers. *Phys. Med. Biol.* **58**, 6065–6077 (2013).
- Kamada, K. *et al.* Human magnetoencephalogram measurements using newly developed compact module of high-sensitivity atomic magnetometer. *Jpn. J. Appl. Phys.* **54**, 026601 (2015).
- Kim, K. *et al.* Multi-channel atomic magnetometer for magnetoencephalography: a configuration study. *Neuroimage* **89**, 143–151 (2014).
- Pfurtscheller, G. & Lopes da Silva, F. H. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* **110**, 1842–1857 (1999).
- Gaetz, W., Macdonald, M., Cheyne, D. & Snead, O. C. Neuromagnetic imaging of movement-related cortical oscillations in children and adults: age predicts post-movement beta rebound. *Neuroimage* **51**, 792–807 (2010).
- Mary, A. *et al.* Aging reduces experience-induced sensorimotor plasticity. a magnetoencephalographic study. *Neuroimage* **104**, 59–68 (2015).
- Robson, S. E. *et al.* Abnormal visuomotor processing in schizophrenia. *Neuroimage Clin.* **12**, 869–878 (2016).
- Uhlhaas, P. J. & Singer, W. Abnormal neural oscillations and synchrony in schizophrenia. *Nat. Rev. Neurosci.* **11**, 100–113 (2010).
- Barratt, E. L. *et al.* Abnormal task driven neural oscillations in multiple sclerosis: A visuomotor MEG study. *Hum. Brain Mapp.* **38**, 2441–2453 (2017).
- Brookes, M. J. *et al.* Optimising experimental design for MEG beamformer imaging. *Neuroimage* **39**, 1788–1802 (2008).

24. Van Veen, B. D., van Drongelen, W., Yuchtman, M. & Suzuki, A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* **44**, 867–880 (1997).
25. Poole, M. & Bowtell, R. Novel gradient coils designed using a boundary element method. *Concepts Magn. Reson. Part B Magn. Reson. Eng.* **31**, 162–175 (2007).
26. Carlson, J. W., Derby, K. A., Hawryszko, K. C. & Weideman, M. Design and evaluation of shielded gradient coils. *Magn. Reson. Med.* **26**, 191–206 (1992).
27. Baillet, S. Magnetoencephalography for brain electrophysiology and imaging. *Nat. Neurosci.* **20**, 327–339 (2017).
28. Brookes, M. J. *et al.* Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc. Natl Acad. Sci. USA* **108**, 16783–16788 (2011).
29. Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M. & Engel, A. K. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* **15**, 884–890 (2012).
30. Jerbi, K. *et al.* Coherent neural representation of hand speed in humans revealed by MEG imaging. *Proc. Natl Acad. Sci. USA* **104**, 7676–7681 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was funded by a Wellcome Collaborative Award in Science (203257/Z/16/Z and 203257/B/16/Z) awarded to G.R.B., R.B. and M.J.B. We also acknowledge the UK Quantum Technology Hub for Sensors and Metrology, funded by the Engineering and Physical Sciences Research Council (EP/M013294/1). We acknowledge Medical Research Council Grants (MR/K005464/1 and MR/M006301/1). The Wellcome Centre for Human Neuroimaging is supported by core funding from Wellcome (203147/Z/16/Z).

OPM sensor development at QuSpin was supported by National Institutes of Health grants R44HD074495 and R44MH110288. The scanner-casts were designed and manufactured by M. Lim at Chalk Studios.

Author Contributions E.B.: system design and fabrication, data collection, data analysis, data interpretation, writing paper. N.H.: system design and fabrication, data collection, data analysis, data interpretation, writing paper. J.L.: system design and fabrication, data collection, data analysis, data interpretation, writing paper. G.R.: system design and fabrication, data collection, data analysis, data interpretation, writing paper. V.S.: system design and fabrication. S.S.M.: data interpretation, writing paper. L.D.M.: data interpretation, writing paper. K.J.M.: data collection, data analysis, data interpretation, writing paper. T.M.T.: data interpretation, writing paper. S.B.: data collection, data interpretation, writing paper. G.R.B.: conceptualization, system design and fabrication, data interpretation, writing paper. R.B.: conceptualization, system design and fabrication, data collection, data analysis, data interpretation, writing paper. M.J.B.: conceptualization, system design and fabrication, data collection, data analysis, data interpretation, writing paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.J.B. (matthew.brookes@nottingham.ac.uk).

Reviewer Information *Nature* thanks S. Baillet and R. Leahy for their contribution to the peer review of this work.

METHODS

OPM-MEG system design and fabrication. *System overview.* An overview of our OPM-MEG system is shown in Extended Data Fig. 7a. The system comprised 13 OPM sensors (QuSpin), which were mounted on the scalp surface (the scalp array), and a further 4 sensors (placed close to the head, but not on the scalp surface) which were used to make reference measurements (the reference array). Each sensor produces an analogue voltage output that is proportional to the magnetic field perpendicular to the laser beam (scaling = 2.7 V nT^{-1}). The sensor outputs were digitized at a sample rate of 1,000 Hz, using a 16-bit digital acquisition (DAQ) system (National Instruments) controlled using custom-written software in LabVIEW. The sensor arrays were housed inside a magnetically shielded room (MSR) to reduce environmental magnetic interference; all control equipment was kept outside the room to minimize its effect on MEG measurements.

Prior to data acquisition, the reference array was used to identify the currents in the coils that best nulled the residual static magnetic field inside the MSR and its dominant first-order spatial variation. Reference sensors were located and oriented such that the three Cartesian components of the magnetic field (B_x , B_y and B_z), and the two dominant field gradients (dB_x/dz and dB_y/dz) could be measured close to the head (Extended Data Fig. 7b). Reference array measurements were input to a proportional integral derivative (PID) algorithm, which was used to control the currents in the field-nulling coils. This allows the calculation of currents which generate fields that are equal and opposite to those measured by the reference array, thus attenuating the static field over the volume spanning the subject's head. This step is key if the head is free to move during MEG data acquisition, as without field nulling, even small changes in head position or orientation (for example, a 4° rotation in a 25-nT field) would produce field variations large enough to render the OPMs inoperable (Fig. 3c).

During data acquisition, all 17 sensors (13 sensor scalp array and 4 sensor reference array) were operated simultaneously, with the reference sensors (which are sufficiently far from the scalp to be insensitive to the neuromagnetic field) used to characterize temporally varying background magnetic interference, which was later regressed from the signals recorded by the scalp-mounted sensors.

The visual cue for paradigm control was controlled by a separate stimulus computer. This was coupled to a data projector that projected the cue image through a waveguide onto a back-projection screen positioned $\sim 40 \text{ cm}$ in front of the subject. The stimulus computer also generated TTL trigger pulses of 1-s duration denoting the onset of the visual cues and the start of the rest periods. These trigger signals were input to the DAQ and sampled, simultaneously with the OPM data, at 1,000 Hz. Throughout data acquisition, an Xbox Kinect camera (Microsoft) was used to measure head movement. Video data were captured using a third computer, which also recorded the trigger channels so that movement data could later be analysed on a trial-by-trial basis.

OPM sensors. The fundamental building block of our system is the OPM sensor. We used compact commercially available sensors^{3,11} (Fig. 1c), allowing a large number to be located flexibly on the scalp surface. Each OPM sensor head contains a semiconductor laser for optical pumping, optics for laser beam conditioning, a $3 \times 3 \times 3\text{-mm}^3$ ^{87}Rb vapour cell and a silicon photodiode for beam detection. The sensor head connects to a small electronics controller via a 5-m cable that is passed through a waveguide in the MSR. The sensor includes three on-board coils, which can be used to null static field components in the cell. Field changes (for example, due to neural currents in the brain) can be detected via the change in transmitted light intensity that they produce (Fig. 1b). The transmitted intensity shows a zero-field resonance, which is a Lorentzian function of the magnetic field components transverse to the laser beam, with a full width at half maximum of around 30 nT. For continuous field measurements, a sinusoidally modulated magnetic field of 1-kHz frequency is applied perpendicular to the laser beam using the on-sensor coils. The depth of modulation of the transmitted light, which is monitored using a lock-in process, is sensitive to the magnitude of the field component along the modulation axis^{7,8}. The amplitude of the two field components perpendicular to the beam can be measured simultaneously by applying oscillating currents to two coils in quadrature. However, here only the radial field component was measured. **Scanner-cast design and 3D printing.** Pre-recorded MRI and cryogenic MEG data were used to inform the design and construction of the individualized scanner-cast. To ascertain the shape of the subject's head, an anatomical MRI scan was acquired using a Philips 3T Ingenia MR system running a T_1 -weighted gradient echo imaging sequence, with an isotropic resolution of 1 mm and a high bandwidth to limit field-inhomogeneity-related distortion. A 3D mesh, representing the outer surface of the head and face, was extracted from this image and then used to define a custom-fitted helmet, in which the OPM sensors could be mounted¹². In a complete system, a large number of sensors would be spaced equidistantly across the entire scalp surface, giving whole brain coverage, but in our prototype system (used for the finger abduction and ball game paradigms), only 13 sensors were available and these had to be positioned for optimal coverage of the sensorimotor

region. To determine the optimal OPM sensor positions, we carried out the visually cued finger abduction experiment on the same subject using a cryogenic MEG system, and localized the source position and orientation using a beamformer. Having computed the location and orientation of the dipolar source, we used a forward model to compute the radial magnetic fields at the scalp surface generated by this dipole. The positions of the scalp-array sensors were then chosen to sample these fields optimally¹². We assessed the effect of crosstalk between sensors, which may occur as a result of constructive interference of fields from adjacent sensors. For the sensor array used here we found these effects to be less than 3% when taking into account the radial field (Extended Data Fig. 8 and Supplementary Information section 8). See Supplementary Information section 7 for details of the scanner-cast that accommodated the 26 sensors that were used in a functional connectivity demonstration (Extended Data Fig. 6).

Field-nulling coils. The remnant Earth's field at the centre of our MSR is $\sim 25 \text{ nT}$, and also shows significant spatial variation, with a gradient of $\sim 10 \text{ nT m}^{-1}$. This means that translation through the inhomogeneous remnant field, or rotation with respect to the field vector, produces field changes at the sensors that are much larger than the neuromagnetic fields of interest (Fig. 3). Given the size of the remnant field and the narrow operational range of the OPMs ($\pm 1.5 \text{ nT}$), such movement is likely to take the sensors outside their operational range. For example, a rotation of less than 4° in a field of 25 nT can produce a change of more than 1.5 nT in the amplitude of the magnetic field along an OPM sensor's sensitive axis, thus rendering it inoperable. The use of the field-nulling coils to reduce the remnant static (Earth's) field over a large central volume inside the MSR is therefore critical if the sensor array is to be operated without being rigidly fixed in position with respect to the MSR.

Here, the remnant field was nulled using a set of coils positioned on either side of the subject (see Fig. 3a). Five different bi-planar coils were constructed independently to null B_x , B_y and B_z and the two dominant field gradients dB_x/dz and dB_y/dz . In contrast to the tri-axial Helmholtz coil arrangement that is commonly used to null the remnant field inside an MSR, the bi-planar coil array allows cancellation of spatially varying fields and also does not significantly limit access to the subject, since the coil windings are confined to two vertical planes (rather than the three pairs of orthogonal planes that would enclose the subject when using three orthogonal Helmholtz coils). The bi-planar coil system therefore offers considerable advantages by eliminating the spatially uniform remnant field and its first-order spatial variation, while maintaining access to the subject.

Magnetic fields from bi-planar coils³¹ were optimized for homogeneity using a harmonic minimization approach²⁶. The current distribution \mathbf{J} was confined to the two planes at $z = \pm a$ in the region $|x|, |y| < L$, and defined using the stream function S as $\mathbf{J} = \nabla S \times \hat{\mathbf{z}}$. S is parameterized as a two-dimensional Fourier series so that

$$S = \sum_n \left[\alpha_n \frac{\cos\left(n - \frac{1}{2}\right)\pi x}{L} + \beta_n \sin \frac{n\pi x}{L} \right] \times \sum_m \left[\gamma_m \frac{\cos\left(m - \frac{1}{2}\right)\pi y}{L} + \delta_m \sin \frac{m\pi y}{L} \right]$$

Optimal values of the Fourier coefficients (α_n , β_n , γ_m , δ_m) were identified by exploiting the symmetry of the field distribution and then minimizing:

$$\sum_t [\mathbf{B}(\mathbf{r}_t) - \mathbf{B}_{\text{target}}(\mathbf{r}_t)]^2 + \omega P$$

Where \mathbf{r}_t characterizes locations within the volume of interest at which a homogeneous field or field gradient (described by $\mathbf{B}_{\text{target}}$) is required, and P represents the power dissipated in the coil. The coefficient ω can be used to adjust the weighting of the power term. Increasing ω reduces the complexity of the wire paths to be fabricated. $\mathbf{B}(\mathbf{r}_t)$ was calculated using the relationship

$$\tilde{\mathbf{B}} = \mu_0 \left\{ [ik_x \hat{\mathbf{x}} + ik_y \hat{\mathbf{y}}] \frac{\sinh(k_r z) - k_r \hat{\mathbf{z}} \cosh(k_r z)}{\cosh(k_r z)} \right\} \tilde{S} e^{-k_r a}$$

where $\tilde{\mathbf{B}}$ and \tilde{S} are the two-dimensional Fourier transforms of the field and stream function with respect to $x(k_x)$ and $y(k_y)$, and $k_r = \sqrt{k_x^2 + k_y^2}$. The upper and lower \sinh and \cosh terms correspond to the case of the stream function having the same or opposite sign on the two planes. Coils were designed to produce a homogeneous (within $\pm 5\%$) field or field gradient over a $0.4 \times 0.4 \times 0.4\text{-m}^3$ central region, which is large enough to span the range of sensor positions during experiments in which head movement is allowed. The wire paths for each coil span an area of $1.6 \times 1.6 \text{ m}^2$ and are layered on two planes of medium-density fibreboard separated by 1.5 m. The coil wire paths and contours of the magnetic field or field gradient in a central plane between the two coils are shown in Extended Data Fig. 9 for each coil. The

field variation was calculated by applying the elemental Biot–Savart law to the digitized wirepaths (see Supplementary Information section 5 for further discussion).

Motion detection and quantification. To quantify head movement, we used a simple optical tracking system based upon a Microsoft Kinect V1-2010 camera which was placed ~1 m in front of the subject. This camera provides a simultaneous 8-bit RGB video stream (640 × 480 pixel display, with 57° horizontal and 43° vertical field of view) and an 11-bit depth image, reconstructed from an infrared projected field. Data from both streams were captured using the MATLAB image acquisition toolbox, at 12 frames per second.

A motion-tracking algorithm was used to track the positions of three black dots on the outer surface of the white scanner-cast. The algorithm was initiated via manual selection of the three dots in the first frame of the video. A threshold detect routine then identified the dots and their centres of mass in all subsequent frames. Pixel numbers were converted to absolute locations in three dimensions by integrating the video data with the depth field. In this way, we defined parameters describing the motion of the three markers throughout the experiment. To quantify motion, movement parameters for all three orientations and all three fiducial locations were concatenated. Figure 2 (left) shows these data as a histogram plotted on a logarithmic scale.

Experimental method. Finger abduction paradigm. Experiments were carried out on a single subject (female, right-handed, age 27), who provided written informed consent (both to participation in the experiments and to release of photographs). The study was approved by the University of Nottingham Medical School ethics committee. The subject undertook multiple repeats of a visuomotor task²⁰. The paradigm comprised visual stimulation with a centrally presented, maximum-contrast, vertical square-wave grating (3 cycles per degree). The grating subtended a visual angle of 8° and was displayed along with a fixation cross on a grey background. In a single trial, the grating was presented for 1 s followed by a 3-s baseline period during which only the fixation cross was shown. During presentation, the participant was instructed to make repeated abductions of their left (non-dominant) index finger. Fifty of these trials were recorded in each experiment. Blocks of ten trials were interspersed with blocks of ten ‘rest’ trials. In these rest trials (also 4 s in duration) the fixation cross remained on the screen, but no finger abduction was made. Averaging across the 50 ‘real’ and 50 ‘rest’ trials independently allowed assessment of SIR (see below). This experiment was repeated six times with the subject stationary and six times with the subject making natural movements. It was also repeated six times using a cryogenic MEG system.

The ‘ball game’ paradigm. The subject was seated in the OPM-MEG system and asked to play a simple ball game in which they continually bounced a table tennis ball on a bat. Ten-second bursts of ‘ping-pong’ were interspersed with 10 s of baseline activity (during which the subject was told to simply hold the bat and ball on their knee). This gave a total trial length of 20 s, and the subject repeated this 29 times. Owing to the location of the OPM sensors in the scanner-cast, the experiment was undertaken with the subject’s non-dominant hand. Trials were cued by a second experimenter who was positioned inside the MSR throughout the experiment and gave verbal instructions to cue the game (they shouted either ‘ping-pong’ or ‘rest’ to begin or stop the game). This experiment was undertaken once in the OPM-MEG system by three subjects.

Cryogenic MEG system data acquisition. To compare OPM-MEG to cryogenic MEG recordings, we carried out the finger abduction experiment, in the same subject, using a 275-channel CTF MEG system (MISL) operating in third-order synthetic gradiometer configuration³². Data were acquired at a sampling frequency of 600 Hz and the subject was seated. Three electromagnetic position indicator coils were placed on the head as fiducial markers (at the nasion, left preauricular and right preauricular points). The locations of these fiducials were tracked continuously during the recording by sequentially energizing each coil and performing a magnetic dipole fit to these data. This allowed both continuous assessment of head movement throughout the measurement, and knowledge of the location of the head relative to the MEG sensors. Prior to the MEG recording, a 3D digitization of the subject’s head-shape, relative to the fiducial markers, was acquired using a 3D digitizer (Polhemus). Co-registration of the MEG sensor geometry to the anatomical MR image was achieved by fitting the digitized head surface to the equivalent head surface extracted from the anatomical image. The subject undertook the experiment six times and a different head digitization was acquired each time.

Data processing: interference rejection. Following data collection, OPM-MEG data were processed to remove magnetic interference. The reference array sensors were located close enough to the scalp array to capture similar environmental interference, but sufficiently far away to be insensitive to the neuromagnetic fields of interest. This meant that the scalp and reference arrays could be used to synthesize a ‘gradiometer’ measurement whereby the scalp array data are de-noised

via regression of the reference array signals¹². The reference array data from all four channels were combined in a single (design) matrix and a regression was used to remove any linear combination of reference array (interference) signals from the scalp array (neuromagnetic) signals. We applied this correction to the raw (that is, unfiltered) data.

Data processing: beamforming. Following interference rejection, all MEG data (OPM and cryogenic) were processed in the same way. Data were initially inspected visually, and any trials with excessive interference were discarded. This resulted in the loss of only one trial (from a cryogenic recording). A beamformer adaptive spatial filtering approach was then employed.

Using a beamformer^{24,33}, an estimate of electrical source strength $\hat{Q}_\theta(t)$ made at time t and a predetermined location in the brain is given by a weighted sum of sensor measurements such that

$$\hat{Q}_\theta(t) = \mathbf{W}_\theta^T \mathbf{m}(t) \quad (1)$$

Here $\mathbf{m}(t)$ is a vector of magnetic field measurements made at time t across all sensors (that is, 13 in our OPM system or 275 in the cryogenic system) and \mathbf{W}_θ is a vector of weighting parameters tuned to a predefined location in source-space and current dipole orientation, represented by the vector θ . The superscript T indicates a matrix transpose.

The weighting parameters are derived on the basis of power minimization. The overall power in the output signal $\hat{Q}_\theta(t)$ is minimized with the linear constraint that power originating from the location and orientation of interest (θ) should remain. Mathematically, the beamformer problem can be written as:

$$\min_{\mathbf{W}_\theta} [\hat{Q}_\theta^2] \text{ subject to } \mathbf{W}_\theta^T \mathbf{L}_\theta = 1 \quad (2)$$

where \hat{Q}_θ^2 represents the source power, given by $\hat{Q}_\theta^2 = \mathbf{W}_\theta^T \mathbf{C} \mathbf{W}_\theta$, \mathbf{C} represents the channel level data covariance matrix calculated over a time-frequency window of interest, and \mathbf{L}_θ is the lead field vector, which is a vector containing a model of the magnetic fields that would be measured at each of the sensors in response to a source of unit amplitude with location and orientation θ . The (regularized) solution to equation (2) is found analytically to be:

$$\mathbf{W}_\theta^T = [\mathbf{L}_\theta^T (\mathbf{C} + \mu \mathbf{\Sigma})^{-1} \mathbf{L}_\theta]^{-1} \mathbf{L}_\theta^T (\mathbf{C} + \mu \mathbf{\Sigma})^{-1} \quad (3)$$

$\mathbf{\Sigma}$ is a diagonal matrix representing the white noise at each of the MEG channels (which we approximate as the identity matrix) and μ is a regularization parameter³⁴. Note that, in addition to source localization, the power minimization term has the desirable effect of reducing artefacts from, for example, muscles (Extended Data Fig. 10 and Supplementary Information section 9).

We sought to examine beta band effects, and so the beamformer weights were constructed, using equation (3), with covariance matrix \mathbf{C} computed using beta band (13–30 Hz) data over a time window spanning the entire experiment²³ (400 s). In all cases (OPM and cryogenic MEG data) the regularization parameter was optimized to give the best SIR (defined as the standard deviation of the finger abduction trials divided by the standard deviation of the rest trials). The lead field was calculated using the analytical formulation first described by Sarvas³⁵. Two other covariance matrices were constructed: \mathbf{C}_a was defined as the data covariance in an ‘active’ window. This spanned the 0 s < t < 1 s window (relative to trial onset) in the case of the finger abduction paradigm, and the 1 s < t < 9 s window (relative to trial onset) in the ball game paradigm. \mathbf{C}_c represented data covariance in a ‘control’ window (1 s < t < 2 s for the finger abduction and 11 s < t < 19 s for the ball game). We then defined a pseudo- T -statistical contrast as

$$T_\theta = \frac{\mathbf{W}_\theta^T \mathbf{C}_a \mathbf{W}_\theta - \mathbf{W}_\theta^T \mathbf{C}_c \mathbf{W}_\theta}{2 \mathbf{W}_\theta^T \mathbf{C}_c \mathbf{W}_\theta}$$

Pseudo- T -statistics were computed at the vertices of a regular 4-mm grid spanning the entire brain. For each voxel, the orientation of each source was based on a non-linear search for the orientation that gave the maximum signal-to-noise ratio³³. This method allowed construction of 3D images showing the spatial signature of maximum change in beta power. These images, averaged across experiments, are shown in Fig. 2 (middle) and in Fig. 4b.

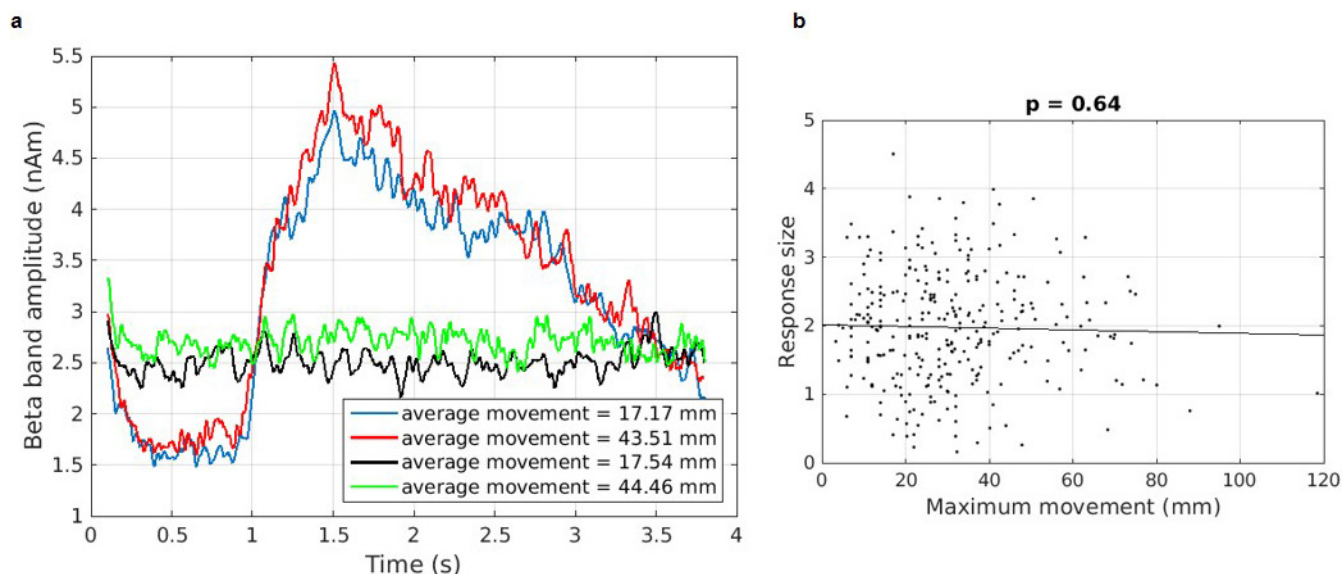
Finally, to investigate the spectrotemporal signature of electrophysiological activity at the location of peak change, a time–frequency spectrogram (TFS) was derived. Here, the peak location/orientation, θ_{peak} , was extracted from the beamformer images, and a time course of electrophysiological activity for that location derived using equation (1) (the data covariance for the weights calculation was expanded to a broad (1–150 Hz) frequency range and covered the full 400 s of data collection). The resulting data were frequency filtered into 31 overlapping frequency bands, and a Hilbert transform was used to generate the amplitude

envelope of oscillations within each band. These envelope time courses were then averaged across trials, and experiments, and concatenated in the frequency dimension to generate a single TFS (shown in Fig. 2 (right) and 4c). The same method was used (with beta band filtered covariance for weights calculation) to examine beta band fluctuations.

Code availability. The MATLAB code used to analyse the current study is available from the corresponding author on request.

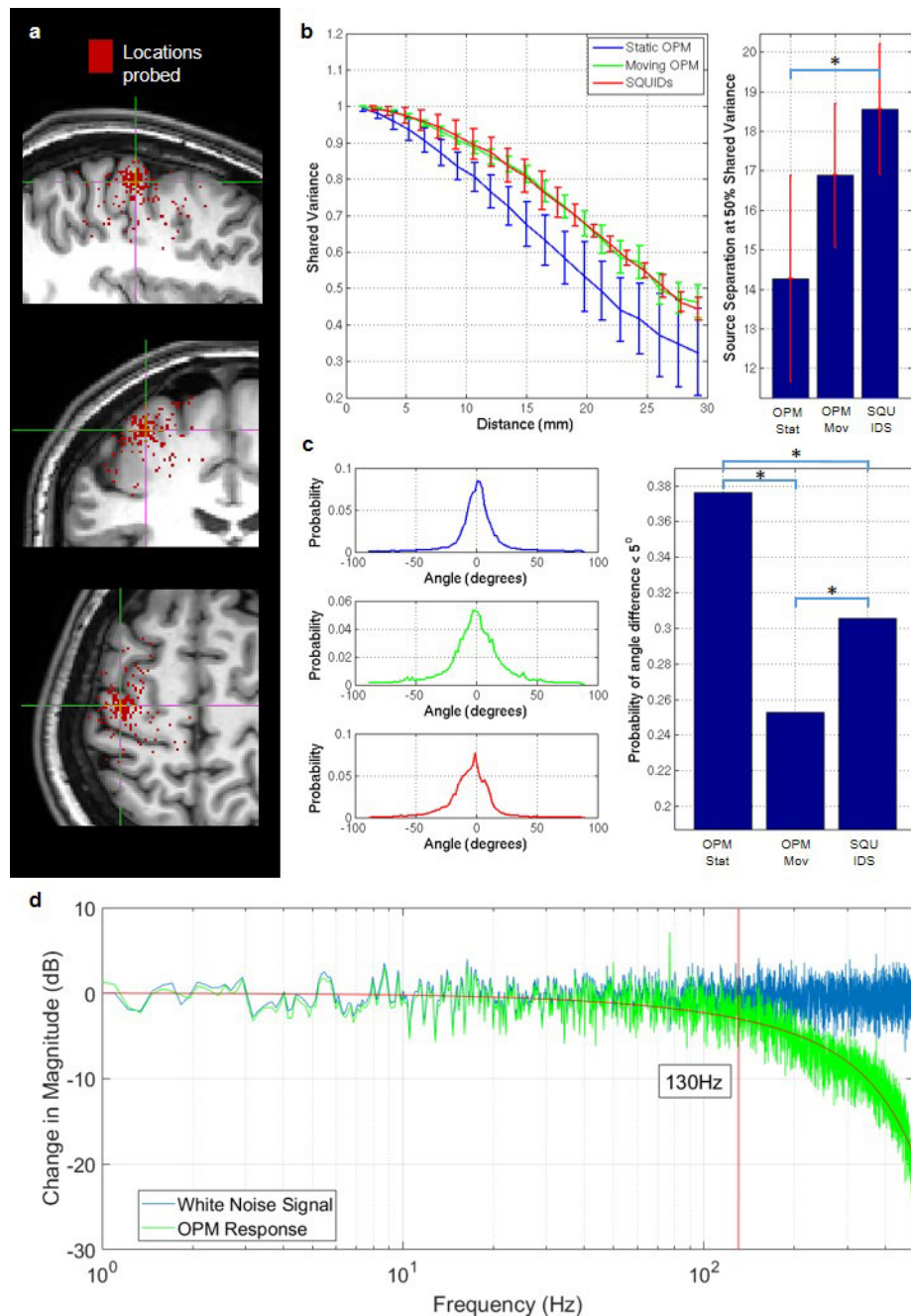
Data availability. The datasets generated during and/or analysed during the current study are available from the corresponding author on request.

31. Yoda, K. Analytical design method of shelf shielded planar coils. *J. Appl. Phys.* **67**, 4349 (1990).
32. Vrba, J. & Robinson, S. E. Signal processing in magnetoencephalography. *Methods* **25**, 249–271 (2001).
33. Robinson, S. & Vrba, J. in *Recent Advances in Biomagnetism* (eds Yoshimoto, T. *et al.*) 302–305 (Tohoku Univ. Press, 1998).
34. Backus, G. E. & Gilbert, F. The resolving power of gross Earth data. *Geophys. J. R. Astron. Soc.* **16**, 169–205 (1968).
35. Sarvas, J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* **32**, 11–22 (1987).



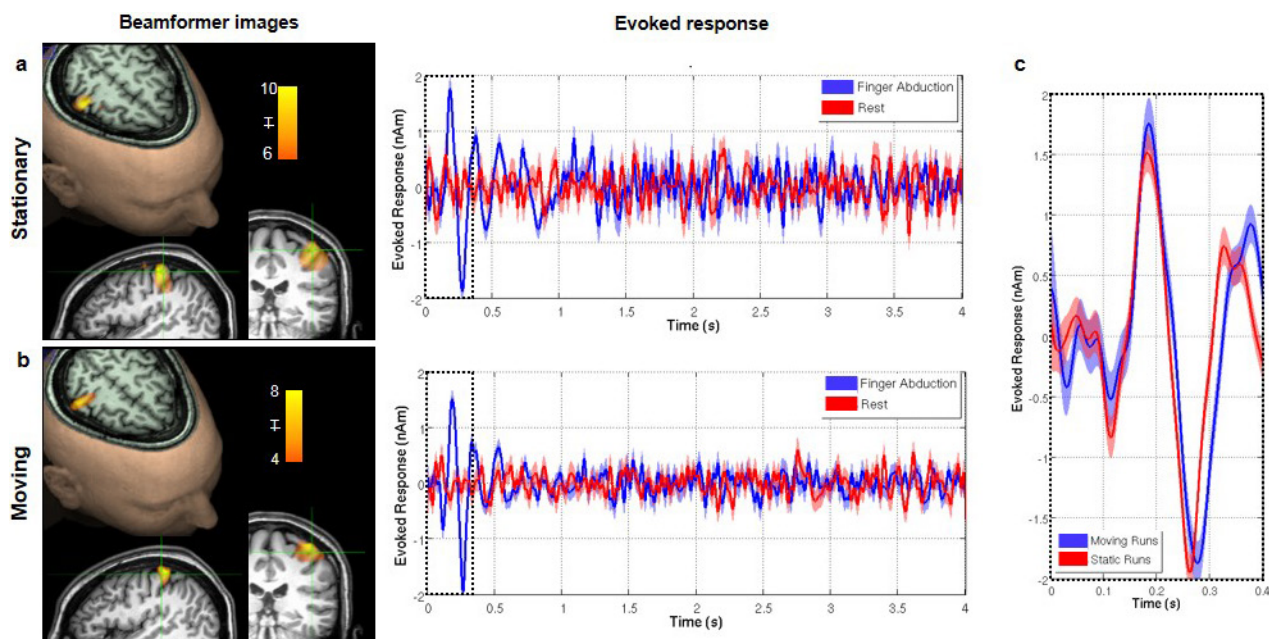
Extended Data Figure 1 | Response magnitude as a function of subject movement. **a**, Beta envelopes for finger abduction trials (blue/red) and resting trials (black/green) in the presence of large movement (red/green) and small movements (blue/black). **b**, The response size (that is, the difference between the mean amplitude during the desynchronization and rebound periods) shown as a function of maximum movement during

a trial. Note that no measurable relationship was found. A significant ($P = 0.0052$, Pearson correlation) baseline shift was observed; this is likely to be a consequence of artefacts in the data generated by electrical activity in muscles controlling the naturalistic movements. See also Supplementary Information section 1.



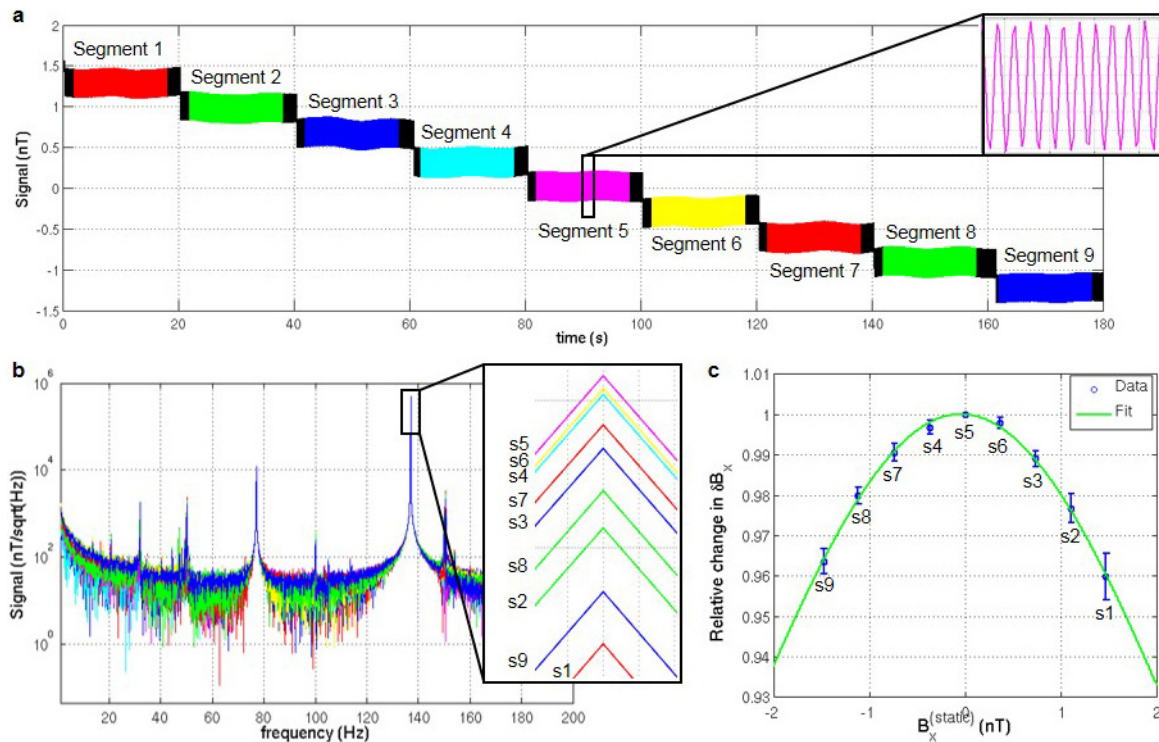
Extended Data Figure 2 | Quantification of spatial and temporal resolution. **a**, A 'seed' location was selected in sensorimotor cortex (at the cross-hairs). Four thousand random 'test' locations (red squares), within 3 cm of the seed, were selected randomly and probed. Shared variance was measured between electrophysiological time courses at the seed and test locations. **b**, Left, correlation between the seed and test time courses plotted as a function of spatial separation. Right, source separation at which shared variance dropped to below 50%. In both cases the error bars show s.d. over test locations. This serves as an absolute quantification of spatial resolution. Note that the OPM array, when static, significantly outperforms SQUIDs ($P = 0.002$, Wilcoxon sum-rank test). **c**, Quantification of the robustness of the source orientation estimation. While source power can vary between experiments, source orientation relies only on the local orientation of the cortical sheet, and should therefore be the same across equivalent experiments. Here the histograms show the source orientation difference (as across runs for 4,000 locations of interest. Note for all three experiments (static OPMs (top), moving OPMs (middle) and SQUIDs (bottom)) the probability distribution peaks at zero as would be expected. The bar chart shows the probability of observing an angular discrepancy below 5°; note

that the OPM array, when static, significantly ($P < 0.05$, permutation test) outperformed the SQUID array in terms of orientation robustness. Moving OPMs demonstrated the lowest robustness; however, this would be expected since the execution of natural movements differed across runs and therefore brain activity in the sensorimotor strip will also differ. The improvement in spatial specificity and robustness in our OPM-MEG system compared to a cryogenic (SQUID) system is likely to be a consequence of two factors: first, the closer proximity of the OPM sensors to the scalp provides higher signal-to-noise ratio in OPMs compared to SQUIDs; second, the scanner-cast ensures that, for each run, OPM sensors are in exactly the same location with respect to the brain. Cryogenic MEG, on the other hand, is subject to co-registration errors. **d**, Quantification of the OPM sensor's frequency response, which also defines its temporal resolution. An OPM sensor was placed in a Helmholtz coil and a white noise current source applied to the coil. The blue line shows the Fourier spectrum of the current source, the green line shows the spectrum of the measured field. Note that sensitivity falls by 3 dB at 130 Hz, giving a temporal resolution of 7.7 ms. See also Supplementary Information section 2.



Extended Data Figure 3 | Evoked response analysis. **a**, Results when the subject was asked to remain still. **b**, Results when the subject was moving. Left panels show functional image: the overlay shows the spatial signature of the 2–30 Hz component of the evoked response, overlaid onto axial, sagittal and coronal slices of the anatomical MRI. Right panels show the time course of the evoked response; finger abduction trials in blue, rest

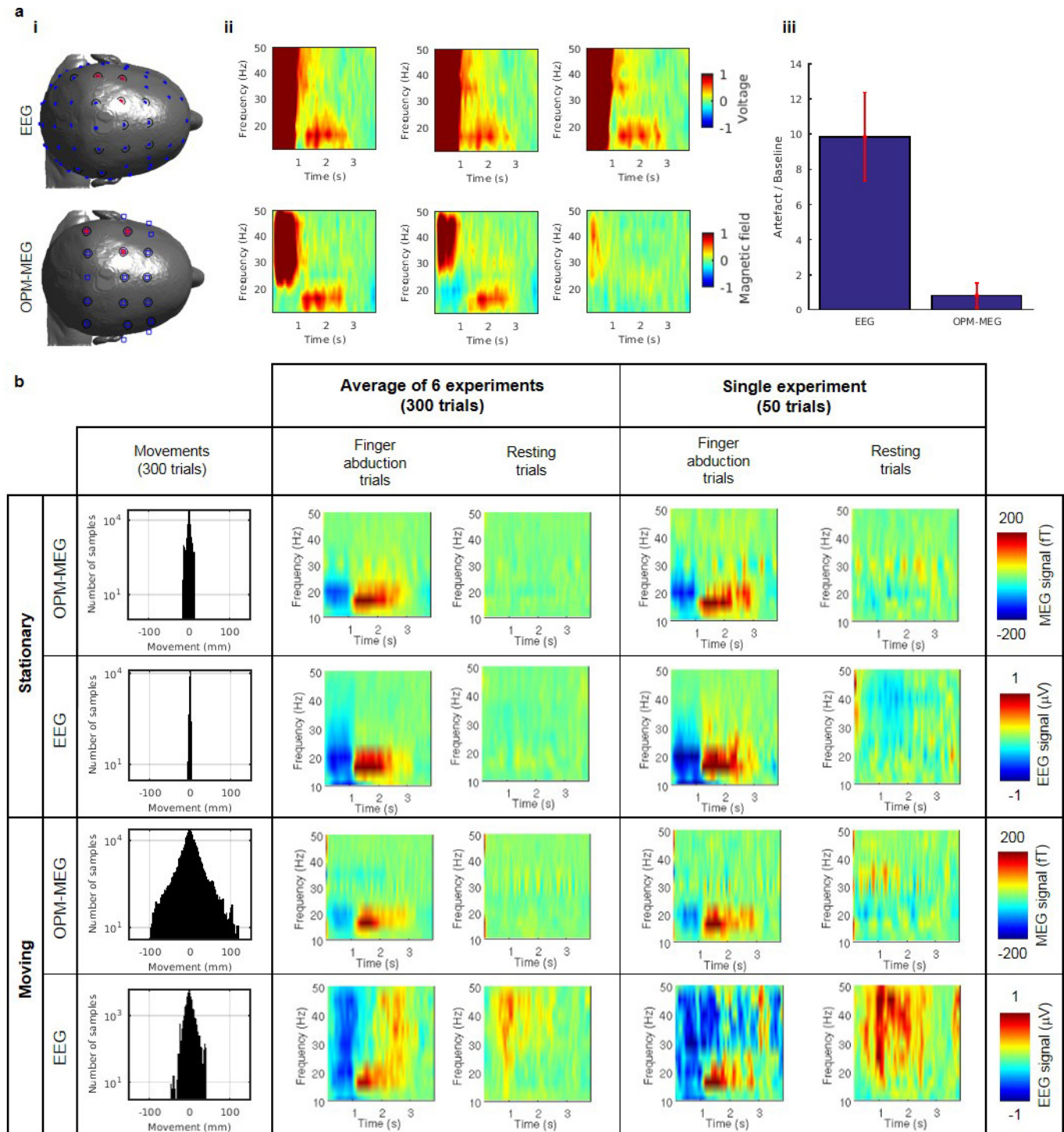
trials in red. The shaded area shows standard error across six experiments. **c**, Direct comparison of the evoked response when the subject was asked to remain still (red) and was moving (blue). No significant SIR difference was found between static and moving runs ($P=0.24$, two-sided Wilcoxon sum-rank test). See also Supplementary Information section 3.



Extended Data Figure 4 | Gain changes with static magnetic field.

a, Raw OPM-MEG data recorded from a single channel during the gain experiment. Data were divided into nine segments (colour-coded here) corresponding to nine different static background magnetic fields ($B_x^{(static)}$), ranging between -1.5 nT and 1.5 nT. The inset plot shows the small oscillating field (δB_x), applied (in this sensor) at 137 Hz using the radially oriented on-sensor coil, which mimics neuromagnetic activity. **b**, Fourier

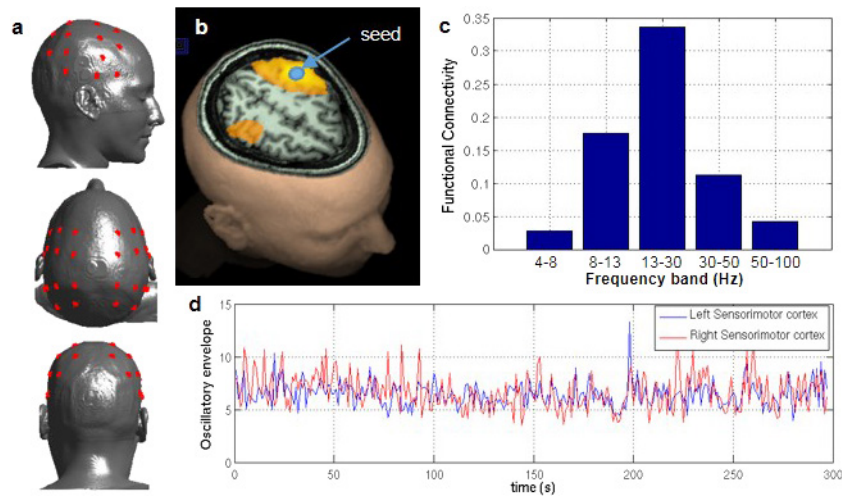
transforms of each data segment. The inset figure shows the height of the 137 -Hz peak for different segments. Note that the peak height changes as a function of static magnetic field. **c**, Fractional change in δB_x as a function of background field $B_x^{(static)}$. The blue circles show the measured data with the standard deviation over the six sensors. The green line shows a fitted Lorentzian function. See also Supplementary Information section 4.



Extended Data Figure 5 | Comparison of EEG and OPM-MEG.

a, Muscle tensing experiment. **i**, Channel montages for EEG (top) and OPM-MEG (bottom). Blue circles show EEG channels used; blue squares show MEG channels used; red stars denote channels used to create **(ii)**; black circles indicate channels used for averages in **(iii)**. **ii**, Time-frequency spectra showing fractional change in oscillatory amplitude, relative to baseline. The three plots show three separate channels, with the muscle artefact visible in the 0–1-s window, when jaw clenching took place. **iii**, Quantitative analysis of the magnitude of the artefact, which was measured to be about ten times larger in EEG. Error bars show s.d. across sensors. **b**, Finger abduction experiment. The four rows show OPM-MEG and EEG data with the subject stationary, followed by

OPM-MEG and EEG data with the subject making natural movements. The left-hand column shows movement parameters. The left and left-centre time–frequency spectra show absolute difference from baseline of the MEG (in fT) and EEG (in μV) signals for individual channels, in the finger abduction and resting trials, respectively. These results have been averaged across all six experiments in both modalities. The right and right-centre time–frequency spectra show equivalent visualizations for a single representative experiment. Notice that, with the head stationary, MEG and EEG show similar results. However with the head moving, EEG data suffer from artefacts generated by muscle activity, to which the MEG data are less susceptible. See also Supplementary Information section 6.

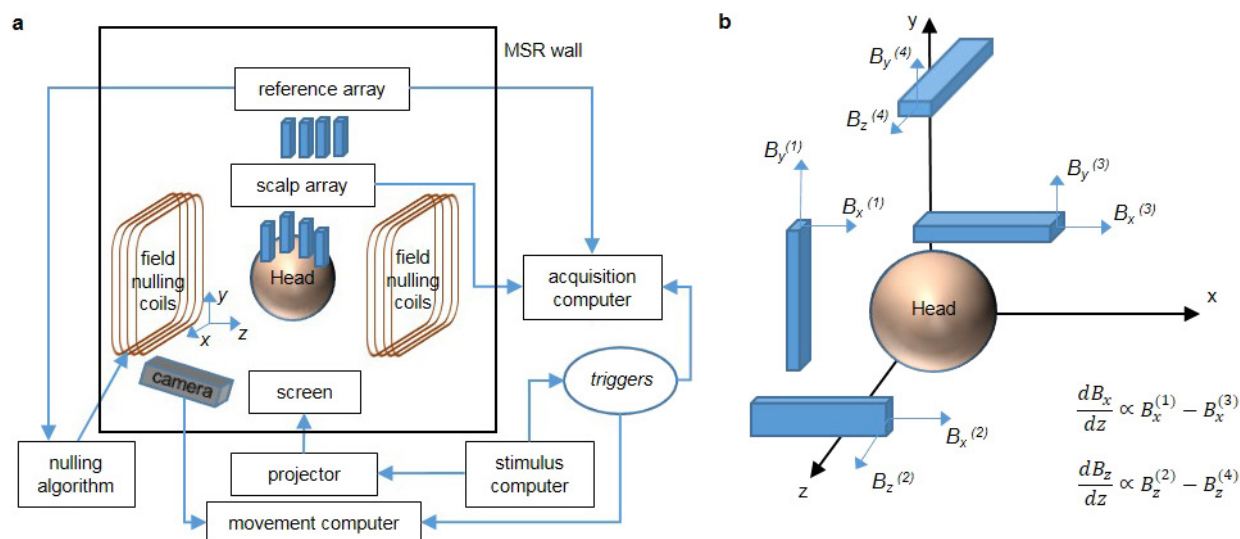


Extended Data Figure 6 | OPM-MEG derived functional connectivity.

A single subject took part in an experiment in which 5 min of OPM-MEG data were acquired in the resting state (subject was told to ‘think of nothing’). The experiment was repeated twice and the results averaged.

a, A 26-channel OPM scalp array, with OPM sensors positioned (using a scanner-cast) approximately to cover the left and right parietal lobes (red circles). MEG data were reconstructed in source space using a beamformer, on a 4-mm grid covering the entire brain. A seed location was selected in left sensorimotor cortex and functional connectivity between the seed and the rest of the brain computed using an amplitude envelope correlation measurement, with correction for signal leakage by

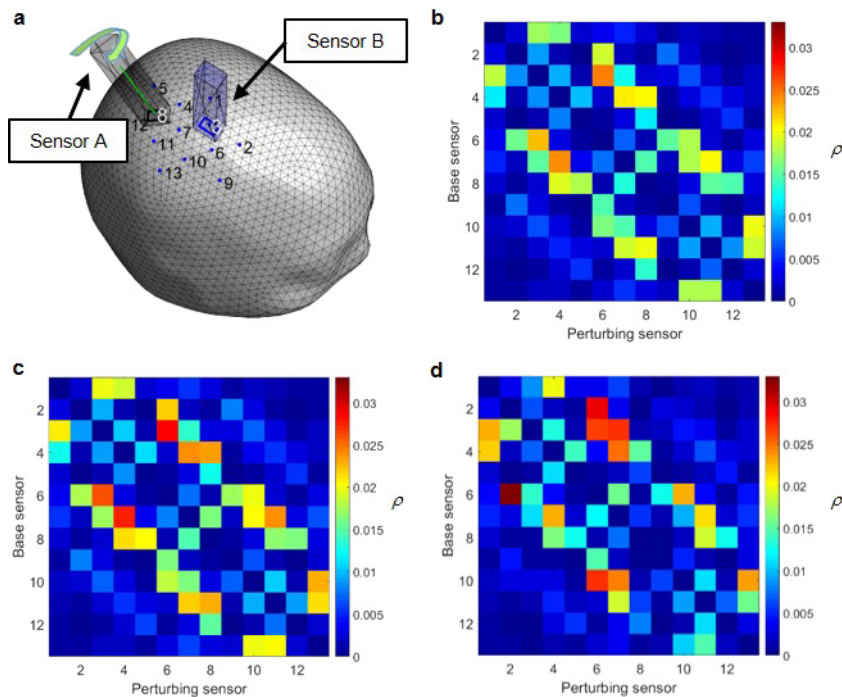
regression. **b**, Regions exhibiting the strongest functional connectivity to the seed location (in the beta frequency band). Note that, in addition to a region around the seed, functional connectivity is observed in the homologous regions of the opposite hemisphere. This reflects long-range functional connectivity within the sensorimotor network. **c**, Functional connectivity strength between left and right primary sensorimotor cortex, plotted as a function of frequency. Note that, as expected, functional connectivity between these regions is greatest in the beta band (13–30 Hz). **d**, An example of beta band envelopes from the left (blue) and right (red) sensorimotor cortices, derived from resting state data. See also Supplementary Information section 7.



Extended Data Figure 7 | An overview of the OPM-MEG system.

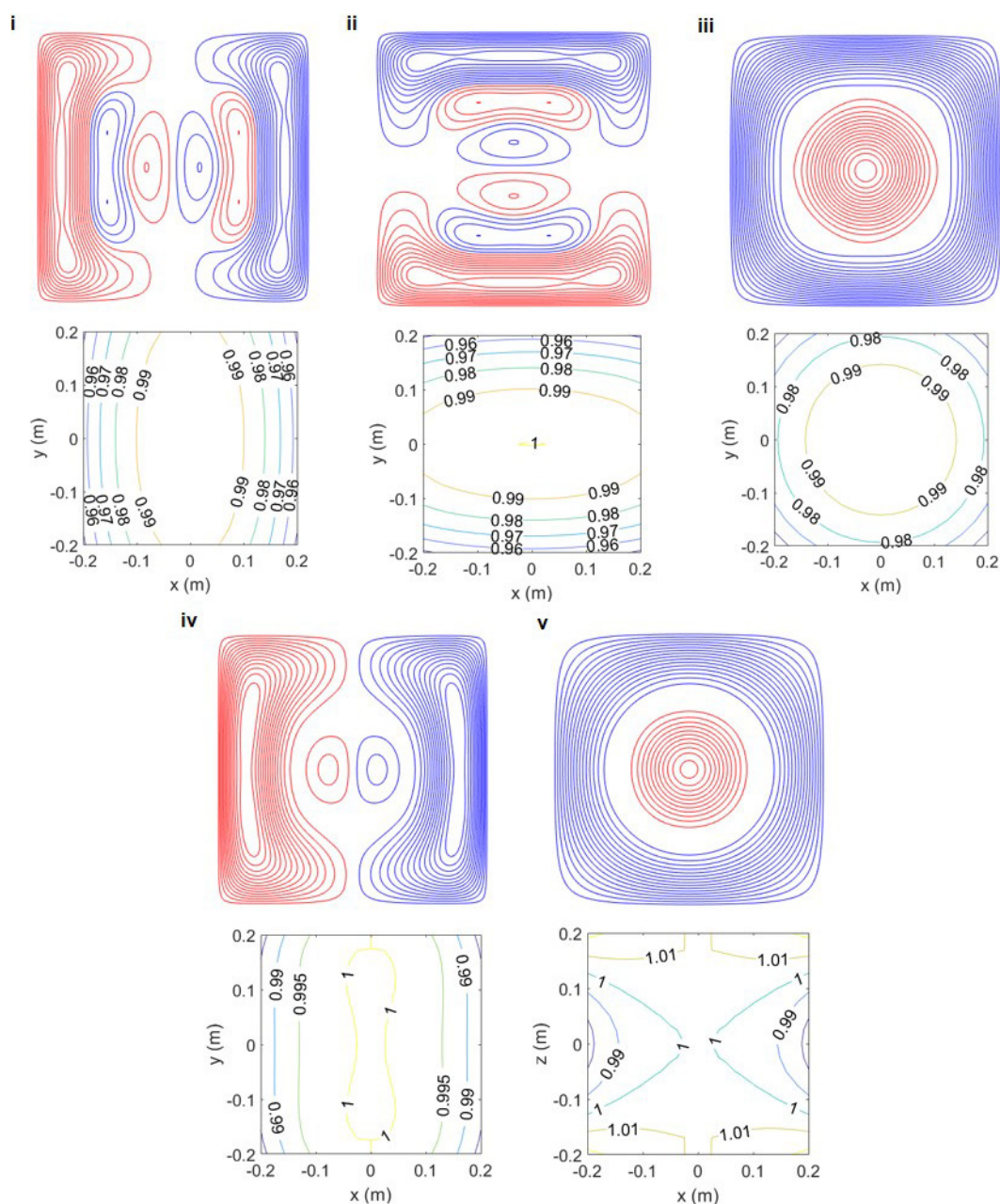
a, Schematic showing an overview of system hardware. **b**, Positioning of the reference sensors relative to the head to allow measurement of the three Cartesian components of the magnetic field, and the two dominant spatial gradients of the field. Each sensor provides measurements of two

components of the magnetic field that are perpendicular to the beam axis. Both components were measured for field nulling, but during experimental measurements only the component of the field along the long-axis of the sensor was measured.



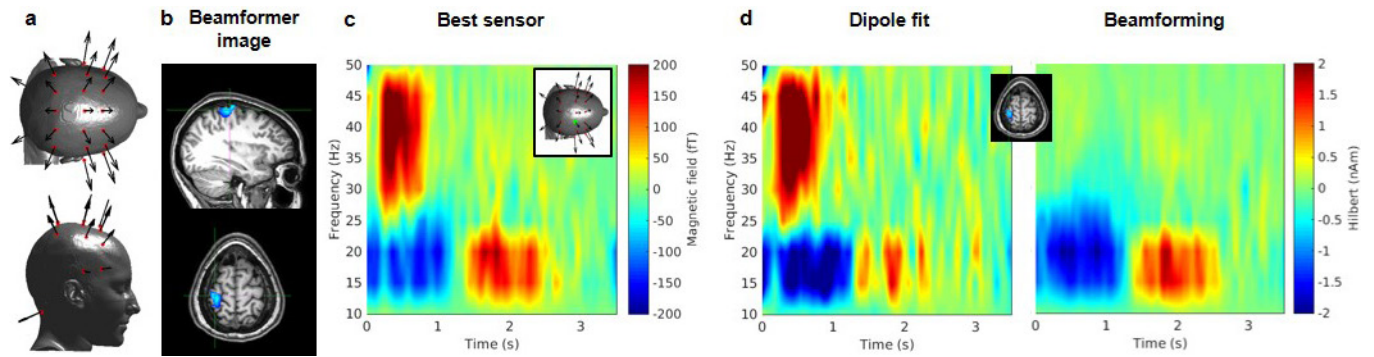
Extended Data Figure 8 | Crosstalk characterization across an OPM array. **a**, Schematic 3D representation of the crosstalk simulation. The head surface is shown with two example sensors. The locations of all 13 sensors are also indicated. We sought to characterize crosstalk between all pairs of sensors in the array. **b**, **c**, Simulated crosstalk between sensors

measured as the ratio of fields generated by the perturbing sensor and the base sensor at the position of the base sensor. This ratio is a periodic function of sensor rotation about the radial orientation; the minimum interaction is shown in **b**, the maximum is shown in **c**. **d**, Experimentally measured crosstalk matrix. See also Supplementary Information section 8.



Extended Data Figure 9 | Coil designs. Wire paths and field plots are shown for the five coils: i, B_x ; ii, B_y ; iii, B_z ; iv, dB_x/dz ; v, dB_z/dz . The upper portion of each part shows the wire paths for one ($1.6 \times 1.6\text{-m}^2$) plane of the bi-planar coil. Red and blue colours indicate clockwise and anticlockwise circulation of the current. The lower portion shows

contours of the field or field gradient strength over the $0.4 \times 0.4\text{-m}^2$ x - y plane located at the centre of the volume of interest ($z=0$). For v, contours are shown in the x - z plane at $y=0$. The field or gradient values are normalized to the value at $x=y=z=0$. Variation from the ideal field distribution is less than 5% over a $0.4 \times 0.4 \times 0.4\text{-m}^3$ central volume.



Extended Data Figure 10 | Removal of muscle artefacts via beamforming. **a**, The montage of OPM-MEG channels used to measure muscle artefact data. **b**, A beamformer image, highlighting a location of interest in right sensorimotor cortex. **c**, The time–frequency response for the best OPM-MEG sensor. **d**, Reconstructed responses from the

over-regularized beamformer (which is analogous to dipole fitting) (left) and unregularized beamformer (right). Note that for unregularized beamforming the muscle artefact is suppressed effectively. See also Supplementary Information section 9.

A TRP channel trio mediates acute noxious heat sensing

Ine Vandewauw^{1,2}, Katrien De Clercq^{1,2,3}, Marie Mulier^{1,2}, Katharina Held^{1,2,3}, Silvia Pinto^{1,2}, Nele Van Ranst^{1,2}, Andrei Segal^{1,2}, Thierry Voet⁴, Rudi Vennekens^{1,2}, Katharina Zimmermann⁵, Joris Vriens^{3§} & Thomas Voets^{1,2§}

Acute pain represents a crucial alarm signal to protect us from injury¹. Whereas the nociceptive neurons that convey pain signals were described more than a century ago², the molecular sensors that detect noxious thermal or mechanical insults have yet to be fully identified^{3–6}. Here we show that acute noxious heat sensing in mice depends on a triad of transient receptor potential (TRP) ion channels: TRPM3, TRPV1, and TRPA1. We found that robust somatosensory heat responsiveness at the cellular and behavioural levels is observed only if at least one of these TRP channels is functional. However, combined genetic or pharmacological elimination of all three channels largely and selectively prevents heat responses in both isolated sensory neurons and rapidly firing C and A δ sensory nerve fibres that innervate the skin. Strikingly, *Trpv1*^{−/−}*Trpm3*^{−/−}*Trpa1*^{−/−} triple knockout (TKO) mice lack the acute withdrawal response to noxious heat that is necessary to avoid burn injury, while showing normal nociceptive responses to cold or mechanical stimuli and a preserved preference for moderate temperatures. These findings indicate that the initiation of the acute heat-evoked pain response in sensory nerve endings relies on three functionally redundant TRP channels, representing a fault-tolerant mechanism to avoid burn injury.

Nociceptors involved in acute heat sensing have cell bodies in the dorsal root and trigeminal ganglia and extend their sensory endings into the skin, mucosa and internal organs^{4,6}. In these nociceptor endings, a rise in temperature induces the firing of action potentials, which propagate towards the dorsal horn or sensory nucleus in the brain, where the sensory information is relayed to second-order neurons⁴. Heat-activated nociceptors are characterized by the expression of the heat- and capsaicin-activated TRP channel TRPV1⁷. However, although ablation⁸ or silencing⁹ of TRPV1-expressing neurons in mice result in almost complete absence of heat nociception, *Trpv1*-knockout mice display only minor deficits in acute noxious heat sensing^{10,11}, indicating that heat-sensitive nociceptors express additional molecular heat sensors. Although various other ion channels, including several other TRP channels and calcium-activated chloride channels of the Anoctamin/TMEM16 family^{4,6}, have been put forward as molecular heat sensors, mice deficient in these various channels showed either no or relatively mild deficits in acute heat sensing^{12–15}. Thus, the molecular basis of acute heat sensing has remained unresolved⁴.

The heat-sensitive TRP channel TRPM3 is expressed in a large subset of TRPV1-expressing sensory neurons, and *Trpm3*^{−/−} knockout mice exhibit reduced heat sensitivity at cellular and behavioural levels^{14,16}. To investigate the consequences of combined elimination of TRPV1 and TRPM3, we analysed the heat responsiveness of *Trpm3*^{−/−}*Trpv1*^{−/−} double knockout (DKO^{M3/V1}) mice. Using calcium imaging, we probed the thermal and chemical responsiveness of sensory neurons isolated from the trigeminal ganglia in primary cultures. Whereas responses to the TRPM3 agonist pregnenolone sulfate and to the TRPV1 agonist

capsaicin were largely eliminated, the proportion of heat-responsive neurons was reduced only slightly compared to that of wild-type mice, with approximately 40% of the DKO^{M3/V1} neurons still exhibiting a robust heat-induced calcium response (Fig. 1a; Extended Data Fig. 1a; see Methods for identification of heat responders). We used a skin-nerve preparation to measure action potentials upon stimulation of the receptive fields of single nerve fibres¹⁷, and recorded robust heat-evoked responses in 33% (6/18) of the mechanosensitive C fibres from DKO^{M3/V1} mice, as compared to 49% (17/35) in wild-type mice ($P = 0.38$; Fisher's exact test; Extended Data Fig. 1b). At the behavioural level, DKO^{M3/V1} mice exhibited vigorous heat avoidance responses in the tail-immersion and hot-plate assays, although withdrawal latencies were prolonged compared to the wild type (Extended Data Fig. 1c). Thus, combined elimination of TRPM3 and TRPV1 results in only very mild deficits in heat responsiveness, implying the existence of one or more additional heat sensors in somatosensory neurons.

We observed that most residual heat-sensitive neurons in the DKO^{M3/V1} mice also responded to allyl isothiocyanate (AITC, 50 μ M; Fig. 1a), a pungent chemical that at this concentration causes selective activation of TRPA1¹⁸. Moreover, in wild-type heat-sensitive neurons, we found extensive overlap of functional expression of TRPA1 with both TRPM3 and TRPV1 (Extended Data Fig. 1d). To test whether TRPA1 contributes to residual heat sensitivity in DKO^{M3/V1} mice, we first evaluated the effect of the selective TRPA1 antagonist HC030031, and found that it caused an almost complete loss of heat-induced calcium responses in DKO^{M3/V1} sensory neurons (Fig. 1b, g). Accordingly, we measured heat-induced inward currents in a subset (9 out of 23) of DKO^{M3/V1} sensory neurons; these currents were inhibited by HC030031 and correlated with cellular responsiveness to AITC (Fig. 1c, d). Next, we produced *Trpv1*^{−/−}*Trpm3*^{−/−}*Trpa1*^{−/−} triple knockout (TKO) mice (Extended Data Table 1), and compared them with the different double knockout (DKO) mouse lines for these three TRP channels (DKO^{M3/V1}, DKO^{M3/A1} (*Trpm3*^{−/−}*Trpa1*^{−/−}) and DKO^{V1/A1} (*Trpv1*^{−/−}*Trpa1*^{−/−}) mice). Sensory neurons isolated from TKO mice showed an almost complete loss of heat responses, which contrasts sharply with the presence of more than 40% heat responders in wild-type mice and in the three DKO mouse lines (Fig. 1e–g).

Several lines of evidence indicate that the loss of heat sensitivity in TKO neurons is specifically due to the combined loss of the three TRP channels. Most importantly, reintroduction via transient transfection of TRPV1, TRPM3 or TRPA1 into TKO neurons restored sensitivity to heat and to the respective channel agonists (Fig. 2a–c, f). Conversely, heat responses in wild-type sensory neurons were suppressed by an inhibitor cocktail consisting of the TRPA1 antagonist HC030031, the TRPV1 antagonist AMG9810, and the TRPM3 antagonist isosakuranetin (Fig. 2d–f). Together, these results indicate that TRPA1, TRPV1 and TRPM3 have critical but redundant roles in heat transduction:

¹Laboratory of Ion Channel Research and TRP Research Platform Leuven (TRPLe), Department of Cellular and Molecular Medicine, University of Leuven, Leuven, Belgium. ²VIB Center for Brain & Disease Research, Leuven, Belgium. ³Laboratory of Experimental Gynecology and Obstetrics, Department of Development and Regeneration, University of Leuven, Leuven, Belgium. ⁴Laboratory of Reproductive Genomics, Department of Human Genetics, University of Leuven, Leuven, Belgium. ⁵Department of Anesthesiology, University of Erlangen-Nürnberg, Erlangen, Germany. [§]These authors jointly supervised this work.

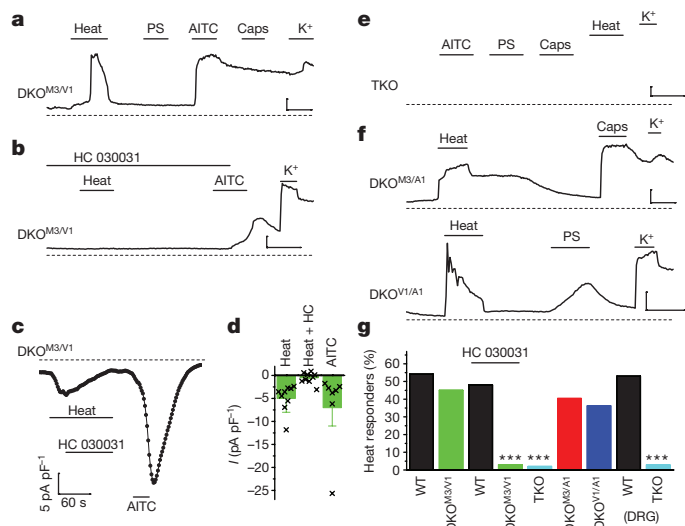


Figure 1 | TRPA1 mediates TRPM3- and TRPV1-independent heat responses. **a, b,** Responses to heat (45 °C) and AITC (50 μ M) in sensory neurons from DKO^{M3/V1} mice in the absence and presence of the TRPA1 antagonist HC030031 (100 μ M). WT, wild-type; PS, pregnenolone sulfate; Caps, capsaicin. **c,** Example of a HC030031-sensitive heat-activated inward current in an AITC-sensitive DKO^{M3/V1} neuron. **d,** Mean (\pm s.e.m.) current densities for experiments as in **c** ($n = 9$ cells). **e, f,** Heat and ligand responses in sensory neurons from TKO and indicated DKO mice. Scale bars in **a, b, e, f,** 60 s/100 nM. **g,** Percentages of heat-responding sensory neurons in mice of the indicated genotypes, in the absence or presence of HC030031. The experiments in **a, b, e, f** and summary data in **g** are representative of the following numbers (n) of tested trigeminal neurons from N mice: wild-type: $n = 296$, $N = 6$; DKO^{M3/A1}: $n = 613$, $N = 5$; DKO^{V1/A1}: $n = 725$, $N = 7$; DKO^{M3/V1}: $n = 1,353$, $N = 9$; TKO: $n = 720$, $N = 11$; WT + HC030031: $n = 208$, $N = 3$; DKO^{M3/V1} + HC030031: $n = 293$, $N = 3$; DRG neurons from wild-type: $n = 133$, $N = 3$; TKO: $n = 181$, $N = 3$. *** $P < 0.00001$; Fisher's exact test with Holm–Bonferroni correction.

functionality of at least one of the three channels is necessary and sufficient to sustain robust heat responses in sensory neurons.

Other aspects of sensory neuron biology were unaltered in the TKO mice. First, functional expression of the cold-activated channel TRPM8 was preserved in the trigeminal ganglia of TKO mice (Extended Data Fig. 2a–c). Compared to wild-type neurons, TKO neurons showed a reduction in the fraction of cold responders (Extended Data Fig. 2c), consistent with the contribution of TRPA1 to cold sensing¹⁹. Second, RNA sequencing (RNA-seq) analysis of isolated sensory ganglia from wild-type, TKO and the different single and double knockout mouse lines for TRPA1, TRPM3 and TRPV1 did not reveal any global changes in mRNA expression profile (Extended Data Fig. 2d). In particular, the mRNA expression levels of other channels implicated in heat sensing, and of a previously defined set of about 200 operational components of sensory neurons²⁰, were not significantly altered in the TKO mice (Extended Data Fig. 2e). Considering that heat-sensitive neurons constitute a large fraction of the total sensory neuronal population in wild-type mice (about 60%; Fig. 1g), these results argue against important alterations at the transcriptional level that would provide an obvious explanation for the striking deficit in heat sensing in the TKO neurons.

A recent study²¹ identified a population of TRPM3- and TRPV1-negative heat-sensitive sensory neurons, and concluded that these neurons rely on TRPM2 for their heat responses. An important piece of evidence to support the involvement of TRPM2 was the finding that TRPM3- and TRPV1-independent heat responses were enhanced by H₂O₂, a known TRPM2 activator. However, as TRPA1 is also sensitive to H₂O₂²², TRPA1 may contribute to the H₂O₂-sensitive heat responses in sensory neurons. Using DKO^{M3/V1} sensory neurons, we confirmed strong potentiation of TRPM3- and TRPV1-independent

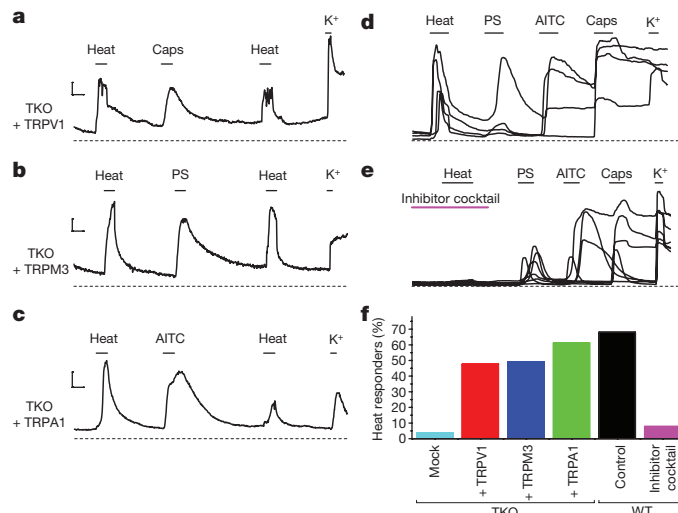


Figure 2 | Properties of TKO sensory neurons. **a–c,** Heat- and ligand-induced calcium responses of TKO trigeminal neurons transfected with TRPV1, TRPM3 or TRPA1. **d, e,** Responses of wild-type trigeminal neurons to heat in the absence (**d**) and presence (**e**) of an inhibitor cocktail of the TRPV1 antagonist AMG9810 (5 μ M), the TRPM3 antagonist isosakuranetin (5 μ M) and the TRPA1 antagonist HC030031 (100 μ M); after the heat stimulus and wash-out of the inhibitor cocktail, responses to the indicated agonists (pregnenolone sulfate, 40 μ M; AITC, 50 μ M; capsaicin, 1 μ M) were conserved. Scale bars in **a–e,** 60 s/100 nM. **f,** Percentages of heat-responding sensory neurons in transfected TKO neurons and in wild-type neurons, in the absence and presence of the inhibitor cocktail. The experiments in **a–e** and summary data in **f** are representative of the following numbers (n) of tested trigeminal neurons from N mice. TKO mock: $n = 277$, $N = 5$; TKO + TRPV1: $n = 130$, $N = 5$; TKO + TRPM3: $n = 189$, $N = 4$; TKO + TRPA1: $n = 289$, $N = 6$; wild-type control: $n = 276$, $N = 6$; wild-type + inhibitor cocktail: $n = 199$, $N = 5$.

heat responses by H₂O₂, including in heat-insensitive neurons, which gained heat sensitivity in the presence of H₂O₂ (Extended Data Fig. 3a, f). However, these potentiated heat responses occurred exclusively in AITC-sensitive neurons, and were fully suppressed in the presence of HC030031 (Extended Data Fig. 3a, b, f). Likewise, H₂O₂ failed to enhance the heat responses in TKO neurons: heat-insensitive TKO neurons never gained heat sensitivity in the presence of H₂O₂ (Extended Data Fig. 3c), and the few residual heat responders (about 3%) among TKO neurons were not potentiated by H₂O₂ (Extended Data Fig. 3d, f). Notably, the residual heat responses in TKO neurons were also not affected by the TRPM2 antagonist 2-APB (Extended Data Fig. 3e). Overall, these data do not provide support for the notion that TRPM2 contributes significantly to heat responses in somatosensory neurons, and instead demonstrate that TRPA1 mediates H₂O₂-sensitive, TRPV1- and TRPM3-independent heat responses in these cells.

Consistent with earlier research²³, we were unable to reliably measure heat responses in non-neuronal CHO cells expressing mouse TRPA1 (Extended Data Fig. 4a, d). However, about 40% of the CHO-TRPA1 cells showed a robust heat-evoked Ca²⁺ response following a brief preincubation with H₂O₂ (Extended Data Fig. 4b, d), and these heat responses were fully suppressed by HC030031 (Extended Data Fig. 4c, d), in line with a recent report showing that the temperature sensitivity of TRPA1 is strongly dependent on the redox state²⁴. Together, our results indicate that the heat responsiveness of mouse TRPA1 is highly dependent on the cellular environment: under control conditions, expression of TRPA1 induces robust heat responses in sensory neurons (Figs 1a, c, d, 2c) but not in non-neuronal cell lines (Extended Data Fig. 4); in both cellular environments, TRPA1-dependent heat responses are strongly enhanced by H₂O₂ (Extended Data Figs 3, 4). In this context, a recent study of TRPA1 in the planarian

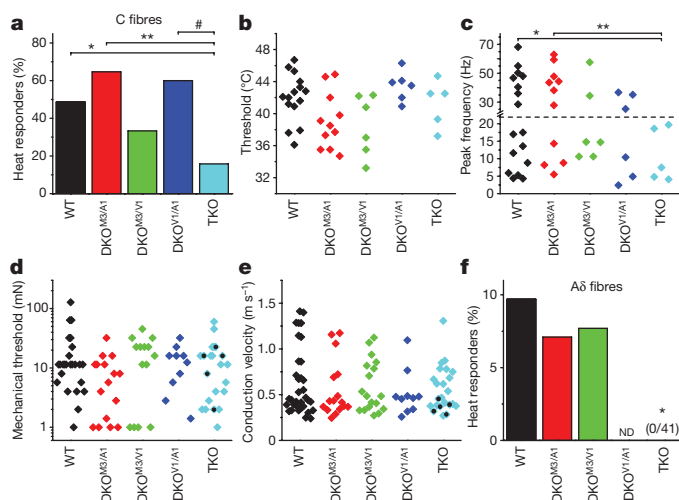


Figure 3 | Heat sensitivity of sensory fibres innervating the skin of TKO mice. **a**, Percentage of mechanosensitive C fibres innervating the paw skin from wild-type, DKO and TKO mice that show a response to heat (48 °C) (* $P=0.041$, ** $P=0.0057$, # $P=0.037$; Fisher's exact test). **b**, Thermal threshold for activation of heat-sensitive C fibres. Horizontal lines in **b**, **d**, **e** indicate mean. **c**, Peak firing frequency of heat-sensitive C fibres during a heating stimulus. The dotted line at 23 Hz indicates the border between low frequency and high frequency firing (see Extended Data Fig. 5). * $P=0.021$, ** $P=0.00086$; one-way ANOVA with Tukey's post-hoc test. **d**, **e**, Mechanosensitivity (**d**) and conduction velocity (**e**) are unaltered in C fibres from the different genotypes. Black dots overlaying TKO data points mark the heat-sensitive C fibres. **f**, Percentage of mechanosensitive Aδ fibres in mice that show a heat response. * $P=0.05$; Fisher's exact test. ND, not done. Number of tested fibres: wild-type C, $n=35$, Aδ, $n=31$; DKO^{M3/A1} C, $n=17$, Aδ, $n=14$; DKO^{M3/V1} C, $n=18$, Aδ, $n=13$; DKO^{V1/A1} C, $n=10$; TKO C, $n=23$, Aδ, $n=41$.

flatworm *Schmidtea mediterranea* (smed-TRPA1) demonstrated that H₂O₂ and reactive oxygen species are rapidly (within seconds) produced as a result of heat exposure, leading to smed-TRPA1 activation and TRPA1-dependent acute heat avoidance²⁵. Our results also do not exclude the possibility that heat-induced TRPA1 activation occurs downstream of an as yet unknown heat sensor.

Next, we evaluated the heat responses of single C and Aδ fibres innervating the skin of wild-type, DKO and TKO mice. The proportion of C fibres that responded to a heat ramp to 50 °C was reduced in TKO mice (Fig. 3a). We further characterized the thermal response characteristics of heat-sensitive C fibres from wild-type, DKO and TKO mice, by analysing their thermal thresholds and peak firing rates (Fig. 3b, c). Overall, the five genotypes did not differ significantly with respect to the thermal threshold ($P=0.07$, one-way ANOVA; Fig. 3b). The peak firing rates of C fibres showed a bimodal distribution, and, accordingly, heat-sensitive C fibres were classified as either high frequency or low frequency, depending on whether their peak firing rate was above or below 23 Hz, respectively (Extended Data Fig. 5). Out of 23 C fibres from TKO mice, none was classified as high-frequency heat-sensitive, which compares to 23% high-frequency heat-sensitive C fibres in wild-type mice (8/39; $P=0.021$, two-tailed Fisher's exact test) and between 11 and 40% in the different DKO mice (Fig. 3c). Notably, TKO C fibres did not differ significantly from wild-type or DKO C fibres with respect to mechanical thresholds and conduction velocities, indicating that the deficit in the TKO C fibres is specific for heat, without affecting general neuronal excitability (Fig. 3d, e). In addition to C fibres, a subset of mouse Aδ fibres also exhibits responses to heat. Whereas approximately 10% of wild-type Aδ fibres respond to the heating ramp, no TKO Aδ fibres out of 41 tested responded to heat (Fig. 3f). Histological analysis did not reveal any significant difference in the nerve bundles innervating the hindpaw skin of TKO mice (Extended Data Fig. 6).

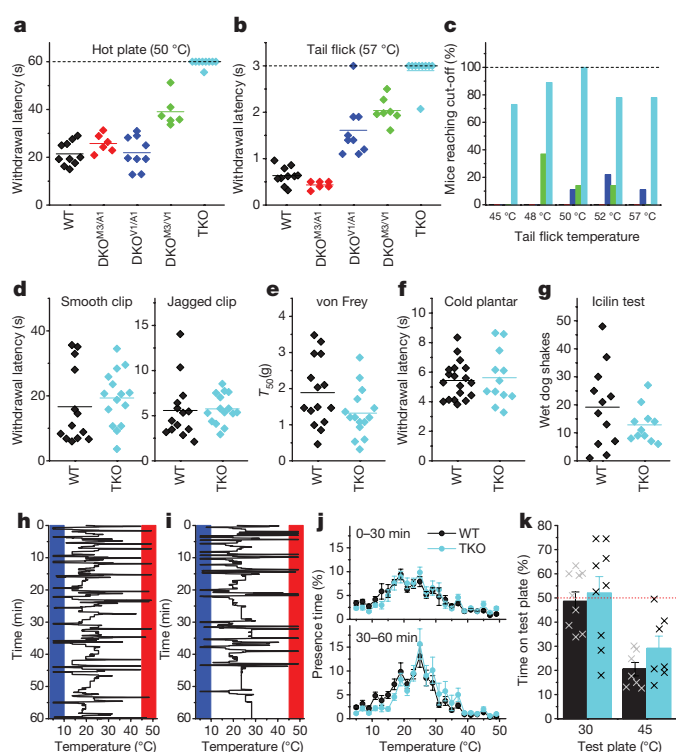


Figure 4 | Selective loss of acute heat nociception in TKO mice.

a, **b**, Withdrawal latencies of male mice in the hot-plate (**a**) and tail-flick (**b**) assays. Dotted lines indicate cut-off times. Results obtained at additional temperatures and statistical analyses are provided in Extended Data Fig. 7. Horizontal lines in **a**, **b**, **d**–**g** indicate mean. **c**, Percentage of tested animals (WT: $N=16$; DKO^{M3/A1}: $N=12$; DKO^{M3/V1}: $N=15$; DKO^{V1/A1}: $N=12$; TKO: $N=16$) that did not withdraw their tail before the cut-off. **d**, Withdrawal latencies of wild-type ($N=13$ – 14) and TKO ($N=15$) mice to smooth plastic and jagged metal clips. $P=0.50$ and 0.86 , respectively; two-sided Student's t -test. **e**, Sensitivity of wild-type ($N=15$) and TKO ($N=15$) mice to calibrated von Frey hairs ($P=0.09$; two-sided Student's t -test). T_{50} , 50% paw withdrawal threshold. **f**, Withdrawal latencies of wild-type ($N=19$) and TKO ($N=12$) mice in the cold plantar test ($P=0.76$; two-sided Student's t -test). **g**, Number of icilin-induced wet dog shakes in wild-type ($N=12$) and TKO ($N=11$) mice ($P=0.20$; two-sided Student's t -test). **h**, **i**, Representative experiments depicting the positions of a wild-type (**h**) and a TKO (**i**) mouse on a thermal gradient. Border zones of the gradient are red (>45 °C) and blue (<10 °C). **j**, Mean presence time \pm s.e.m. in the different temperature zones during the first and second 30-min periods ($N=10$ mice for each genotype). **k**, Two-temperature choice test with control plate set at 30 °C and test plate set at 30 °C or 45 °C. Mean percentage \pm s.e.m. of the total time spent on the test plate ($N=9$ mice for each genotype; $P=0.77$ at 30 °C, $P=0.16$ at 45 °C; two-sided Student's t -test). Both genotypes showed statistically significant avoidance at 45 °C ($P=0.001$; one-sample t -test versus 50%).

Overall, these results indicate that combined ablation of TRPA1, TRPV1 and TRPM3 causes a striking deficit in the heat sensitivity of Aδ and C fibres.

Next, we compared TKO mice with wild-type mice and the different DKO mouse lines with respect to acute heat nociception. In both the tail-immersion and hot-plate assays, the TKO mice showed a severe deficit in heat avoidance behaviour, to such a degree that the majority of TKO animals reached the cut-off times without withdrawal response to a range of noxious temperatures over 45 °C (Fig. 4a–c; Extended Data Figs 7, 8a, b; Supplementary Video 1). In comparison, the different DKO mice all exhibited vigorous withdrawal responses to noxious heat, albeit with overall longer latencies in the case of the DKO^{M3/V1} and DKO^{V1/A1} mice (Fig. 4a–c; Extended Data Fig. 7). Note that, by using a cut-off latency set at four times the mean response latency of wild-type mice, we did not observe superficial signs of acute

skin burns immediately following the hot-plate or tail-immersion experiments in mice of any genotype. However, visual re-inspection of the animals three days after the 57°C tail-immersion experiments revealed severe tissue damage (scarring, necrosis) in the tails of all TKO mice, which was never observed in wild-type mice or any of the DKO mice (Extended Data Fig. 8d–f). Thus, TKO mice lack the acute heat response that is necessary to avoid burn injury.

The absence of a rapid withdrawal response to noxious heat was not due to a general lack of sensitivity to aversive stimuli, as TKO mice showed wild-type-like pain responses in the tail clip assay (Fig. 4d), normal mechanosensitivity of their hindpaws in the von Frey assay (Fig. 4e; Extended Data Fig. 8c), and an unaltered withdrawal latency in the cold plantar assay²⁶ (Fig. 4f). Likewise, the number of icilin-induced wet dog shakes, a measure of TRPM8-dependent cooling-induced behaviour, was not significantly different between wild-type and TKO mice (Fig. 4g).

In addition to reflex avoidance responses to noxious cold or hot temperatures, mice exhibit thermotaxis behaviour towards their preferred ambient temperature, typically in the moderately warm (25–35°C) range. To test whether TKO mice exhibited alterations in thermal preference, mice were allowed to move freely on a 1.2-m-long thermal gradient ranging from 5 to 50°C for 60 min. Wild-type and TKO mice showed extensive exploratory behaviour and exhibited a similar mean distribution on the gradient (Fig. 4h–j). Overall, we did not detect any significant differences between wild-type, TKO and the different DKO mice with respect to the total covered distance on the gradient or the peak occupied temperature (T_{peak} ; Extended Data Fig. 9). Mice of all genotypes explored the full length of the gradient, including the border zones (marked in blue (below 10°C) and red (above 45°C) in Fig. 4h, i), and the combined time spent in these two zones was similar between genotypes (Extended Data Fig. 9c). However, whereas wild-type and the different DKO mice showed a preference for the blue (cold) versus the red (hot) border zone, this preference was not observed in the TKO mice (Extended Data Fig. 9a).

In the two-temperature choice test, where animals were given the choice between a control plate at 30°C and a test plate at 45°C, animals of all genotypes showed a similar preference for the control plate (Fig. 4k; Extended Data Fig. 9c). A striking difference between the genotypes was found when we analysed the timing of individual visits to the test plate: the duration of the longest visit to the 45°C plate was significantly longer for TKO mice than for wild-type or the different DKO mice (Extended Data Fig. 9b). A further experiment with the test plate set at 50°C was prematurely terminated, as we observed that several TKO mice would stay for more than 60 s on the test plate, risking burn injury. Together, these data demonstrate that thermal preference is largely preserved in TKO mice, despite the strong deficit in acute heat-induced pain. These results contrast with the phenotype of mice in which all TRPV1-positive neurons were ablated, which not only lack noxious heat responses but are also indifferent to temperatures between 30 and 50°C in the two-temperature choice assay^{8,27}. The ability of TKO mice to discriminate temperatures in the 30–50°C range may depend on the very small but consistent subset (about 3%) of TRPV1-, TRPM3- and TRPA1-independent heat-responsive sensory neurons (Fig. 1g; Extended Data Fig. 3), or on thermal stimulation of skin keratinocytes that then signal to sensory neurons via diffusible messengers²⁸. The subset of low-frequency heat-sensitive C fibres, which is conserved in the TKO mice (Fig. 3), are potential candidates to convey such keratinocyte-dependent signals. Finally, heat avoidance behaviour may be secondary to changes in core body temperature. In particular, the preoptic area of the hypothalamus contains temperature-sensitive neurons that detect small increases in local brain temperature and initiate both physiological heat loss mechanisms²⁹ and cold-seeking behaviour³⁰.

Although the nociceptive neurons involved in acute pain signalling in mammals were described more than a century ago², the molecular mechanisms whereby these neurons detect harmful signals have

remained largely unresolved. The TKO mouse represents, to our knowledge, the first demonstration in mammals of elimination of the pain response to a noxious physical stimulus at the level of the transducer ion channels. The presence of three redundant molecular heat-sensing mechanisms with overlapping expression in nociceptive neurons represents a powerful fail-safe mechanism that ensures avoidance of noxious heat even under conditions where the function of one or two heat sensors is compromised, for instance due to channel desensitization or naturally occurring inhibitory ligands. We speculate that similar redundancy may operate in mammals for the detection of other noxious physical stimuli.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 August 2017; accepted 16 February 2018.

Published online 14 March 2018.

- Grayson, M. Pain. *Nature* **535**, S1 (2016).
- Sherrington, C. S. *The Integrative Action of the Nervous System* (Oxford Univ. Press, 1906).
- Julius, D. TRP channels and pain. *Annu. Rev. Cell Dev. Biol.* **29**, 355–384 (2013).
- Vriens, J., Nilius, B. & Voets, T. Peripheral thermosensation in mammals. *Nat. Rev. Neurosci.* **15**, 573–589 (2014).
- Geffeney, S. L. & Goodman, M. B. How we feel: ion channel partnerships that detect mechanical inputs and give rise to touch and pain perception. *Neuron* **74**, 609–619 (2012).
- Peirs, C. & Seal, R. P. Neural circuits for pain: recent advances and current views. *Science* **354**, 578–584 (2016).
- Caterina, M. J. et al. The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
- Pogorzala, L. A., Mishra, S. K. & Hoon, M. A. The cellular code for mammalian thermosensation. *J. Neurosci.* **33**, 5533–5541 (2013).
- Binshtok, A. M., Bean, B. P. & Woolf, C. J. Inhibition of nociceptors by TRPV1-mediated entry of impermeant sodium channel blockers. *Nature* **449**, 607–610 (2007).
- Caterina, M. J. et al. Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**, 306–313 (2000).
- Davis, J. B. et al. Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. *Nature* **405**, 183–187 (2000).
- Huang, S. M., Li, X., Yu, Y., Wang, J. & Caterina, M. J. TRPV3 and TRPV4 ion channels are not major contributors to mouse heat sensation. *Mol. Pain* **7**, 37 (2011).
- Park, U. et al. TRP vanilloid 2 knock-out mice are susceptible to perinatal lethality but display normal thermal and mechanical nociception. *J. Neurosci.* **31**, 11425–11436 (2011).
- Vriens, J. et al. TRPM3 is a nociceptive channel involved in the detection of noxious heat. *Neuron* **70**, 482–494 (2011).
- Cho, H. et al. The calcium-activated chloride channel anoctamin 1 acts as a heat sensor in nociceptive neurons. *Nat. Neurosci.* **15**, 1015–1021 (2012).
- Vriens, J. et al. Opening of an alternative ion permeation pathway in a nociceptive TRP channel. *Nat. Chem. Biol.* **10**, 188–195 (2014).
- Zimmermann, K. et al. Phenotyping sensory nerve endings *in vitro* in the mouse. *Nat. Protocols* **4**, 174–196 (2009).
- Everaerts, W. et al. The capsaicin receptor TRPV1 is a crucial mediator of the noxious effects of mustard oil. *Curr. Biol.* **21**, 316–321 (2011).
- Karashima, Y. et al. TRPA1 acts as a cold sensor *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **106**, 1273–1278 (2009).
- Usoskin, D. et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).
- Tan, C. H. & McNaughton, P. A. The TRPM2 ion channel is required for sensitivity to warmth. *Nature* **536**, 460–463 (2016).
- Andersson, D. A., Gentry, C., Moss, S. & Bevan, S. Transient receptor potential A1 is a sensory receptor for multiple products of oxidative stress. *J. Neurosci.* **28**, 2485–2494 (2008).
- Story, G. M. et al. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **112**, 819–829 (2003).
- Moparthy, L. et al. Human TRPA1 is a heat sensor displaying intrinsic U-shaped thermosensitivity. *Sci. Rep.* **6**, 28763 (2016).
- Arenas, O. M. et al. Activation of planarian TRPA1 by reactive oxygen species reveals a conserved mechanism for animal nociception. *Nat. Neurosci.* **20**, 1686–1693 (2017).
- Brenner, D. S., Golden, J. P., Vogt, S. K. & Gereau, R. W., IV. A simple and inexpensive method for determining cold sensitivity and adaptation in mice. *J. Vis. Exp.* (97): (2015). 10.3791/52640
- Mishra, S. K., Tisel, S. M., Orestes, P., Bhargoo, S. K. & Hoon, M. A. TRPV1-lineage neurons are required for thermal sensation. *EMBO J.* **30**, 582–593 (2011).
- Caterina, M. J. & Pang, Z. TRP channels in skin biology and pathophysiology. *Pharmaceuticals (Basel)* **9**, e77 (2016).

29. Song, K. *et al.* The TRPM2 channel is a hypothalamic heat sensor that limits fever and can drive hypothermia. *Science* **353**, 1393–1398 (2016).
30. Tan, C. L. *et al.* Warm-sensitive neurons that control body temperature. *Cell* **167**, 47–59.e15 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all members of the Laboratories of Ion Channel Research and Experimental Gynecology and Obstetrics for comments and discussion, and K. Luyten for assistance with histological and *in situ* staining. This work was supported by grants from the KU Leuven Research Council (PF-TRPLe and C1-TRPLe to Tho.V. and R.V.), the Research Foundation-Flanders (FWO G.084515N to J.V. and Tho.V. and G.099114N to Tho.V. and Thi.V.), the Queen Elisabeth Medical Foundation for Neurosciences (to Tho.V.), the Belgian Foundation Against Cancer (to J.V. and Tho.V.) and the Planckaert-De Waele fund (to J.V.). K.D.C. and K.H. are holders of a doctoral fellowship of the FWO Belgium.

Author Contributions I.V., M.M. and J.V. performed cellular calcium imaging experiments. I.V. designed, performed and analysed all skin nerve recordings,

with technical input and guidance from K.Z. and A.S. K.H., J.V. and Tho.V. performed patch-clamp experiments. I.V., K.D.C., K.H., S.P., M.M. and J.V. performed behavioural experiments. K.D.C. and N.V.R. performed histological analyses. I.V., Thi.V. and A.S. performed and analysed RNA-seq experiments. S.P. generated DKO and TKO mice, with supervision by R.V. I.V. and Tho.V. wrote the manuscript with input from all co-authors. J.V. and Tho.V. initiated and supervised the entire project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to Tho.V. (thomas.voets@kuleuven.vib.be) or J.V. (joris.vriens@kuleuven.be).

Reviewer Information *Nature* thanks J. Wood and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. Wild-type C57BL/6J and *Trpa1*^{-/-}*Trpm3*^{-/-}, *Trpa1*^{-/-}*Trpv1*^{-/-}, *Trpm3*^{-/-}*Trpv1*^{-/-} and *Trpa1*^{-/-}*Trpm3*^{-/-}*Trpv1*^{-/-} mice on a C57BL/6J background, aged 10–14 weeks, were used for all experiments. Male mice were used for behavioural experiments, unless stated otherwise; primary neurons and skin-nerve preparations were obtained from both male and female mice. Mice were housed in a conventional facility at 21 °C on a 12-h light–dark cycle with unrestricted access to food and water. All experiments using animals were approved by the KU Leuven Ethical Committee Laboratory Animals under project numbers P021/2012 and P265/2015.

Isolated sensory neurons. After mice were killed by CO₂ inhalation, trigeminal ganglia were excised, washed in neurobasal A medium (Invitrogen) supplemented with 10% fetal calf serum (basal medium), and then incubated for 45 min at 37 °C in a mix of 1 mg/ml collagenase (Gibco) and 2.5 mg/ml dispase (Gibco). Digested ganglia were gently washed twice with basal medium and mechanically dissociated by passage through syringes fitted with increasing needle gauges. Neurons were seeded on poly-L-ornithine/laminin-coated glass-bottomed chambers (Fluorodish, WPI) and cultured overnight at 37 °C in 5% CO₂ in B27 (Invitrogen) supplemented neurobasal A medium, containing 2 ng/ml GDNF (Invitrogen) and 10 ng/ml NT4 (Peptrotech). For transfection experiments, trigeminal ganglia neurons were grown for 4 days before transfection with 0.5 µg cDNA encoding mouse TRPA1, TRPV1 or TRPM3 in the bicistronic pCAGGS-IRES-GFP vector, using Lipofectamine 2000 (Invitrogen) as a transfection reagent. Neurons were kept at 37 °C in 5% CO₂ for 3 days after transfection before calcium imaging.

Cell culture. CHO cells stably expressing mouse TRPA1 (CHO-A1 cells, kindly provided by A. Patapoutian) were cultured in DMEM containing 10% fetal bovine serum, 2% glutamax (Invitrogen), 1% non-essential amino acids (Invitrogen) and 200 µg/ml penicillin/streptomycin at 37 °C in a humidity-controlled incubator with 5% CO₂²³. The cell line tested negative for mycoplasma contamination. Cell line authentication was not performed, but TRPA1 expression was verified based on AITC responsiveness.

Calcium imaging. Changes in intracellular calcium concentration were monitored using ratiometric Fura-2-based fluorimetry. Cells were loaded with 2 µM Fura-2-acetoxymethyl ester (Alexis Biochemicals) for 30 min at 37 °C. Fluorescence was measured during alternating illumination at 340 and 380 nm using either a Cell^M (Olympus) or Eclipse Ti (Nikon) fluorescence microscopy system, and absolute calcium concentration was calculated from the ratio of the fluorescence signals at these two wavelengths ($R = F_{340}/F_{380}$) as $[Ca^{2+}] = K_m \times (R - R_{min})/(R_{max} - R)$, where K_m , R_{min} and R_{max} were estimated from *in vitro* calibration experiments with known calcium concentrations. The bath solution contained (in mM): 138 NaCl, 5.4 KCl, 2 CaCl₂, 2 MgCl₂, 10 glucose, and 10 HEPES, pH 7.4. Allylisothiocyanate, pregnenolone sulfate, and capsaicin were obtained from Sigma-Aldrich and dissolved in bath solution from a 1,000× stock solution in DMSO. At the end of each experiment, cells were subjected to a depolarizing solution containing 50 mM KCl, and non-responsive cells were excluded from analysis.

The following procedure was used in sensory neurons to distinguish stimulus-induced responses from background variations in calcium concentration or calcium-independent temperature-induced changes in Fura 2 fluorescence. We calculated the time derivatives of the calcium concentration, F_{340} and F_{380} as well as the s.d. of these derivatives in the absence of any stimulus. A positive response was noted when the following criteria were simultaneously met: 1) the stimulus caused an increase of $d[Ca^{2+}]/dt$ exceeding $3 \times$ s.d.; 2) $dF_{340}/dt > 3 \times$ s.d.; and 3) $dF_{380}/dt < 3 \times$ s.d.

For CHO-A1 cells, a positive heat response was defined as a response that exceeded the mean response in the presence of HC030031 by at least five times the s.d., as described elsewhere²⁹.

Whole-cell patch-clamp recordings. Membrane currents in isolated sensory neurons were measured in the whole-cell configuration at a constant holding potential of -40 mV using an EPC-10 amplifier and PatchMaster Software (HEKA Elektronik). The extracellular solution contained (in mM): 140 NaCl, 4 KCl, 2 CaCl₂, 2 MgCl₂, 10 HEPES (pH 7.4 with NaOH), supplemented with 100 nM tetrodotoxin (TTX) to block voltage-gated sodium channels. The internal solution contained (in mM): 140 CsCl, 1 MgCl₂, 1 EGTA, 10 HEPES and 5 tetraethylammonium chloride (pH 7.2 with CsOH).

Skin-nerve recordings. Single-fibre recordings from cutaneous C and Aδ fibres of the saphenous nerve were obtained using an isolated skin-nerve preparation as described previously¹⁷. A skin flap of the hind paw hairy skin and the innervating saphenous nerve were excised and fixed in an organ bath, corium side up. The chamber containing the skin tissue was continuously superfused with pre-warmed (30 °C) synthetic interstitial fluid (SIF) solution, consisting of (in mM) 107.8 NaCl, 3.5 KCl, 26.2 NaHCO₃, 1.67 NaH₂PO₄, 9.64 Na-gluconate, 0.69 MgSO₄, 7.6 sucrose, 5.05 glucose, and 1.53 CaCl₂. The pH was buffered to 7.4 using carbogen

(95% oxygen and 5% carbon dioxide). The nerve trunk was placed on a mirror in a separate chamber filled with paraffin oil, desheathed and divided under microscopic control until a unit with a single distinguishable receptive field was obtained. Initially, mechanosensitive receptive fields were mapped using a blunt plastic rod. Once a distinct receptive field was identified, electrostimulation was applied to the receptive field through a metal microelectrode to determine the fibre's conduction velocity. Values of conduction velocity were used for fibre classification with a cut-off criterion of <1.4 m/s for unmyelinated C fibres. A marking technique, in which latency shifts are provoked through simultaneous application of a mechanical and electrical stimulus to the receptive field, was then applied, ensuring recording from a distinct single fibre. To assess the sensory properties of these mechanically sensitive fibres and for further classification into subclasses, the mechanical, cold and heat sensitivity of the individual fibres was tested. First, the mechanical threshold of the fibre was characterized using calibrated von Frey monofilaments ranging from 1 mN to 128 mN in a geometric scale. Subsequently, a rounded plastic ring (~8 mm diameter) was placed above the receptive field. The volume of fluid within the ring was evacuated and cold (by fluid replacement with cold buffer at 4 °C), followed by radiant heat (ramp of 30–50 °C delivered by a halogen lamp with an 8-mm focus beam) stimuli were applied. Heat thresholds were taken as the temperature at which the second spike occurred. Data were acquired using the Dapsys data acquisition system (B. Turnquist, Bethel University, USA).

Behavioural tests. Considering that we tested animals of five different genotypes, with a priori unknown somatosensory properties, we were not able to perform a reasonable power analysis to determine the sample size for behavioural experiments. Therefore, sample sizes were chosen to be consistent with those reported in similar behavioural studies. In nociceptive assays, mice from different genotypes were tested for all the different stimuli, so no randomization of experimental groups was necessary.

In the tail-immersion assay, mice were immobilized in aluminium foil, which allowed free tail movement. The tip of the tail (one-third of the length) was immersed in a water bath maintained at 45, 48, 50, 52, or 57 °C. The latency to a nociceptive response (withdrawal of the tail) was measured. To prevent acute tissue damage, the tail was removed from the bath immediately after a nociceptive response or upon reaching a cut-off time, which was set at four times the mean response latency of wild-type mice.

In the hot-plate assay, mice were individually confined in a Plexiglas chamber on a metal surface set at 50, 55 or 58 °C and the latency to a nociceptive response (licking or shaking of hind paws, jumping) was measured. To prevent tissue damage, mice were removed from the hot plate after a nociceptive response or upon reaching a cut-off time of three times the mean response latency of wild-type mice.

In the thermotaxis assay, mice were individually tracked for 60 min in a thermal gradient apparatus (Bioseb), which consisted of a controlled and stable temperature gradient from 5 to 50 °C on an aluminium floor. Tracking was performed using a video camera-based system and the time spent in two-degree zones was determined.

In the two-temperature choice assay, mice were placed in a chamber containing two identical, adjacent floor platforms (Bioseb) with one set to 30 °C and the other set to 45 or 50 °C. Mice were free to explore for 10 min, and the total time spent on each surface and the number of times an animal moved to an adjacent platform were analysed.

In the tail clip assay, either a plastic mini-peg (Maped) or metal crocodile clamp (Teishin, Japan), exerting a force of 5 N, was applied to the root of the tail, and the latency to a nociceptive response was measured.

In the Von Frey test, mice were individually placed in a chamber with a wire mesh floor, and allowed to habituate for 15 min before testing. Mechanical responses were tested by stimulating the middle plantar surface of the hind paw with von Frey monofilaments using the up-and-down method, starting with 1 g. Biting, licking, or withdrawal during or immediately following the 3-s stimulus were considered as a positive response.

In the rotarod assay, mice were initially placed on the stationary rod. After habituation, rotation was started at 4 r.p.m. (rotations per minute), accelerating over a 5-min period to 40 r.p.m. This was done in three sessions during four trial days and one test day. The time taken for the mouse to fall from the rod was recorded.

In the elevated maze test, mice were placed in a plus-shaped apparatus with two open and two enclosed arms, connected by a central platform, which was elevated from the floor. Mice were individually placed in the centre of the maze and observed for 5 min. The time spent in the open and enclosed arms was recorded using ANY-maze tracking software (Stoelting).

To evoke icilin-induced wet dog shakes, mice were intraperitoneally injected with 50 mg/kg icilin, which induces robust shivering and shaking in mice. The incidence of these wet dog shakes was counted for 20 min after icilin administration.

In the cold plantar test, mice were acclimated on a 4-mm-thick glass plate, and then the withdrawal latency was measured when a compressed dry ice pellet was held against the glass surface underneath the hindpaw. The latency to withdrawal was tested three times for each hindpaw with an interval of ~15 min, and averaged to obtain a single value per mouse²⁶.

RNA-seq. After mice were killed by CO₂ inhalation, DRGs were isolated and kept in RNAlater Stabilization Reagent (Qiagen) at -20 °C until further processing. Total RNA was extracted using the RNeasy Mini Kit (Qiagen). The sequencing libraries were prepared according to the standard Illumina TruSeq mRNA stranded sample preparation protocol and each library was sequenced on an Illumina HiSeq 2500 according to the manufacturer's recommendations, generating 50-bp single-end reads at the Genomics Core Leuven. Reads were aligned using the open-source, splice-aware tool Tophat (v2.0.13). Read counts per feature (genes or exons) were done by HTSeq (v0.5.3p3). Count data were analysed to identify differentially expressed genes using DESeq2 (v1.01.1) at FDR 10%.

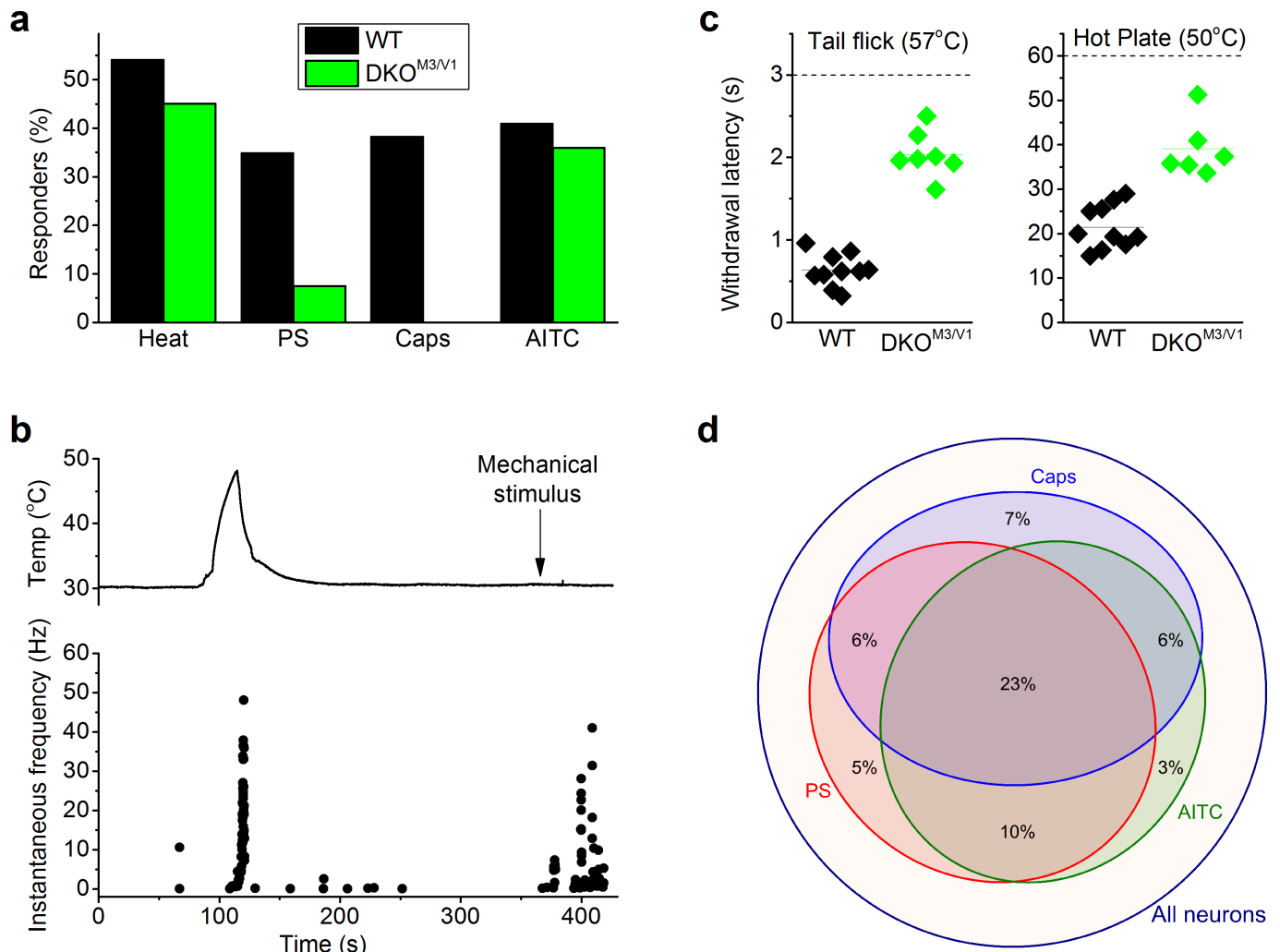
In situ hybridization. Mice were transcardially perfused via the left ventricle with 25 ml PBS followed by 25 ml neutral buffered formalin (10% (w/v), pH 7.4). Subsequently, trigeminal ganglia were dissected and immersion-fixed in neutral buffered formalin for 6 h. Thereafter, trigeminal ganglia were transferred to PBS for 1 h, and then to 70% ethanol until paraffin embedding. Paraffin-embedded tissue was cut into 4-µm-thick sections using an RM2235 rotary microtome (Leica), and probed with digoxigenin-labelled antisense *Trpm8* RNA probes. Antisense RNA probes were generated using the MAXIsript SP6/T7 Transcription Kit (ThermoFisher Scientific), using a 204-bp cDNA fragment of mouse *Trpm8* (accession number NM_134252.3) corresponding to nucleotides 1779–1982 (exons 13 and 14). Hybrid molecules were detected using the TSA Plus Fluorescein Kit (Perkin Elmer) according to the manufacturer's instructions.

Nerve bundle anatomy. The plantar skin of the mouse hind paw was fixed in 4% paraformaldehyde and further processed for staining. Four-micrometre sections

were incubated for 30 min in H₂O₂ (3%) and antigen retrieval was performed in 0.01 M citrate buffer (pH 6, 90 °C, 1 h). Slides were gradually cooled down and incubated in blocking buffer containing 2% bovine serum albumin (Sigma-Aldrich), 1% non-fat dry milk (Nestlé), 0.1% Tween (Merck), complemented with 1:25 normal goat serum (X0907, Dako), before overnight incubation with the primary antibody (4 °C, 1 µg/ml rabbit polyclonal anti-PGP9.5, Z5116, Dako). Slides were washed and blocked again with blocking buffer (15 min), after which they were incubated with the biotinylated swine anti-rabbit secondary antibody (30 min, 1:400; Z0196, Dako) supplemented with normal mouse serum (1:25; Dako). Staining was initiated by the addition of horseradish peroxidase-conjugated streptavidin (30 min, 1:1,000; P0397, Dako) followed by 3-amino-9-ethylcarbazole substrate-chromogen (25 min). Sections were counterstained with Mayer haematoxylin and mounted in glycerin jelly. For each mouse, the nerve bundle areas of 20 sections were evaluated with an average distance of 0.1 mm between the cross sections.

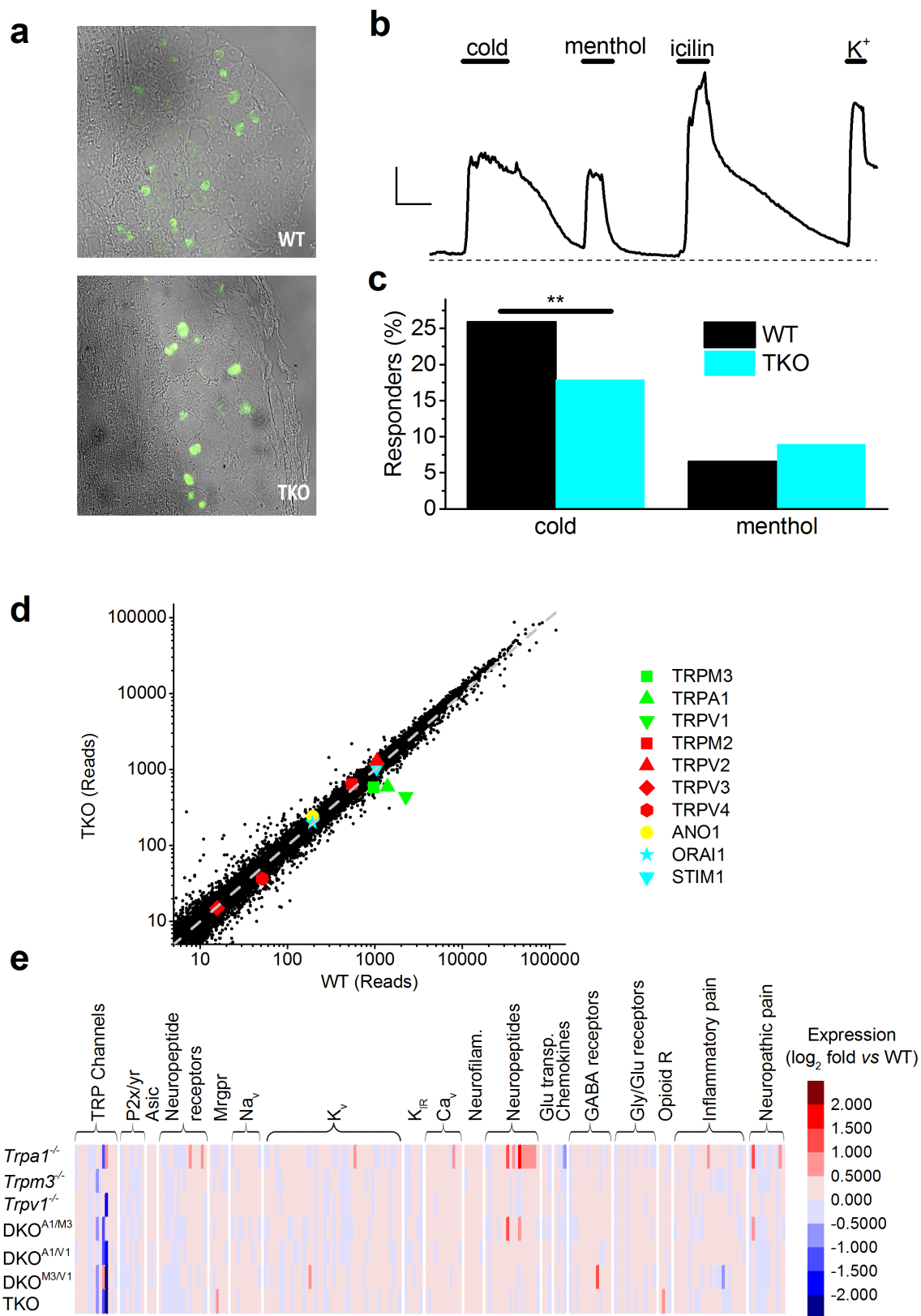
Statistical analysis. Data analysis was performed using Origin software (v8.6-9.0; OriginLab). Group data are represented as mean ± s.e.m. from *n* cells or fibres or *N* animals. Tests used for statistical comparison between groups are indicated in the text and figure legends. Whenever parametric statistics were performed, normality was assessed using the Shapiro–Wilk test and variances were checked to ensure that they were similar between groups being compared. If conditions for parametric tests were not met, non-parametric tests were used. The investigators were not blinded to allocation during experiments. Analysis of data from calcium imaging, skin-nerve and behavioural experiments was performed without knowledge of the genotype of the cells, tissues or mice.

Data availability. Sequencing data that support the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) and are accessible under the BioProject ID PRJNA431738. All other relevant data sets are included within the article or are available from the corresponding authors on reasonable request.



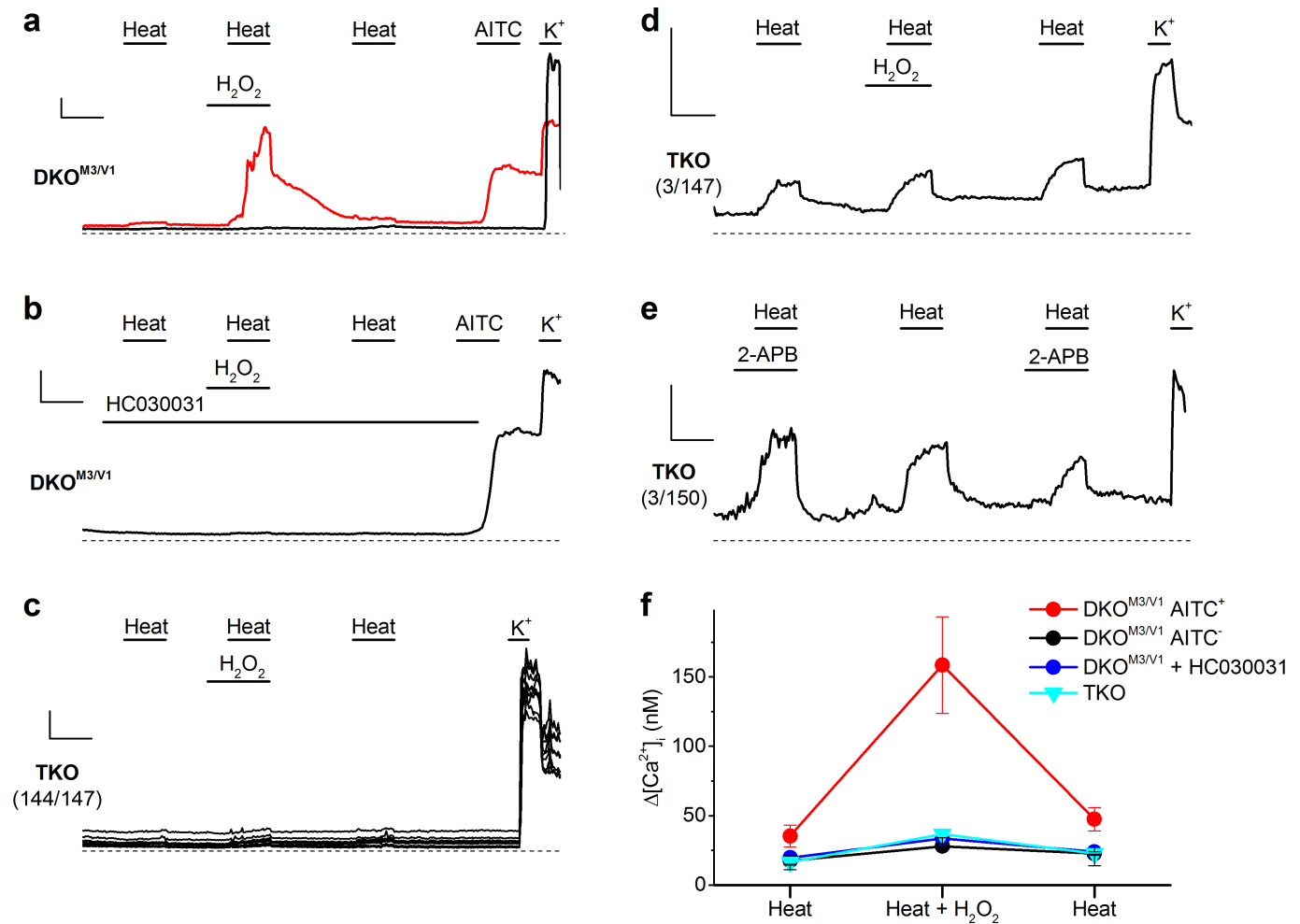
Extended Data Figure 1 | Heat sensitivity is preserved in *Trpm3*^{-/-}*Trpv1*^{-/-} double knockout (DKO^{M3/V1}) mice. **a**, Percentages of wild-type (296 neurons from 6 mice) and DKO^{M3/V1} (1,353 neurons from 9 mice) trigeminal neurons responding to the indicated stimuli. **b**, Single-fibre nerve recording of a mechano- and heat-sensitive C fibre innervating the skin of a DKO^{M3/V1} mouse. **c**, Withdrawal latencies of wild-type

($N = 10$) and DKO^{M3/V1} ($N = 7$) mice in the tail-flick (at 57 °C) and hot-plate (at 50 °C) assays. Dotted lines indicate the cut-off time. Horizontal lines indicate mean. **d**, Venn diagram showing the level of functional coexpression of TRPM3, TRPV1 and TRPA1 in wild-type trigeminal neurons, based on responsiveness to PS, Caps and AITC ($n = 662$ neurons from 6 mice).



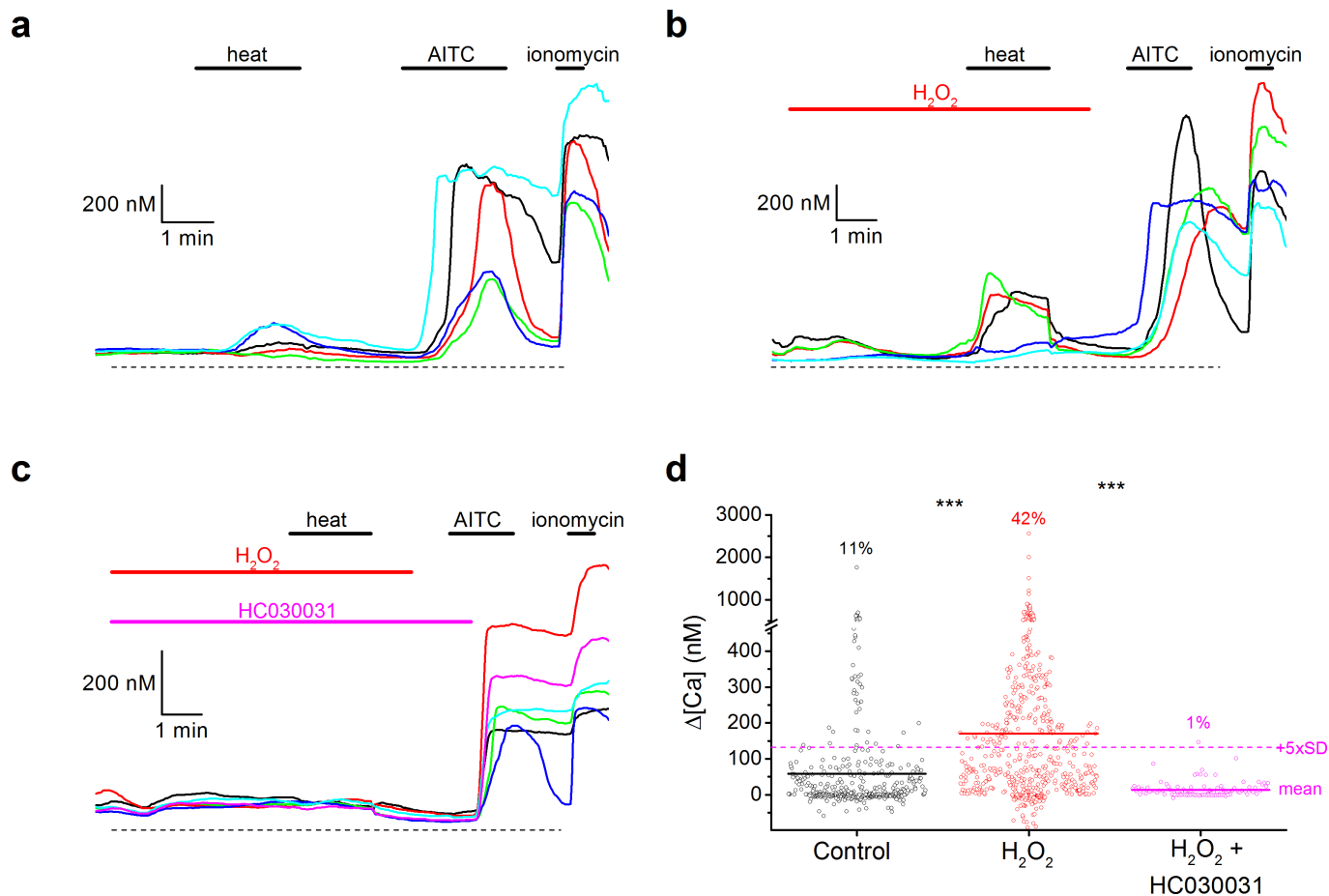
Extended Data Figure 2 | TRPM8 expression and RNA-seq analysis of sensory neurons. **a**, *In situ* hybridization using a TRPM8-specific probe in trigeminal ganglia of wild-type and TKO mice. **b**, Example of a sensory neuron from a TKO mouse responding to cold (15 °C) and the TRPM8 agonists menthol (50 μ M) and icilin (1 μ M). Scale bar, 6s/100 nM. **c**, Percentages of cold-responding and menthol-responding neurons in wild-type ($n = 225$ from 3 mice) and TKO mice ($n = 189$ from 3 mice). ** $P = 0.0089$; Fisher's exact test. **d**, Transcriptome-wide comparison of

mRNA expression between wild-type and TKO mice. Several ion channels that have been previously implicated in heat sensing are indicated. **e**, Heat map showing the differential expression levels in single, double and triple knockout mice of a set of about 200 genes implicated in somatosensation²⁰. Expression levels are defined as the log₂ fold change compared to the wild type. Except for *Trpm3*, *Trpv1* and *Trpa1*, no genes showed more than 50% up- or downregulation of expression in TKO compared to wild-type mice.



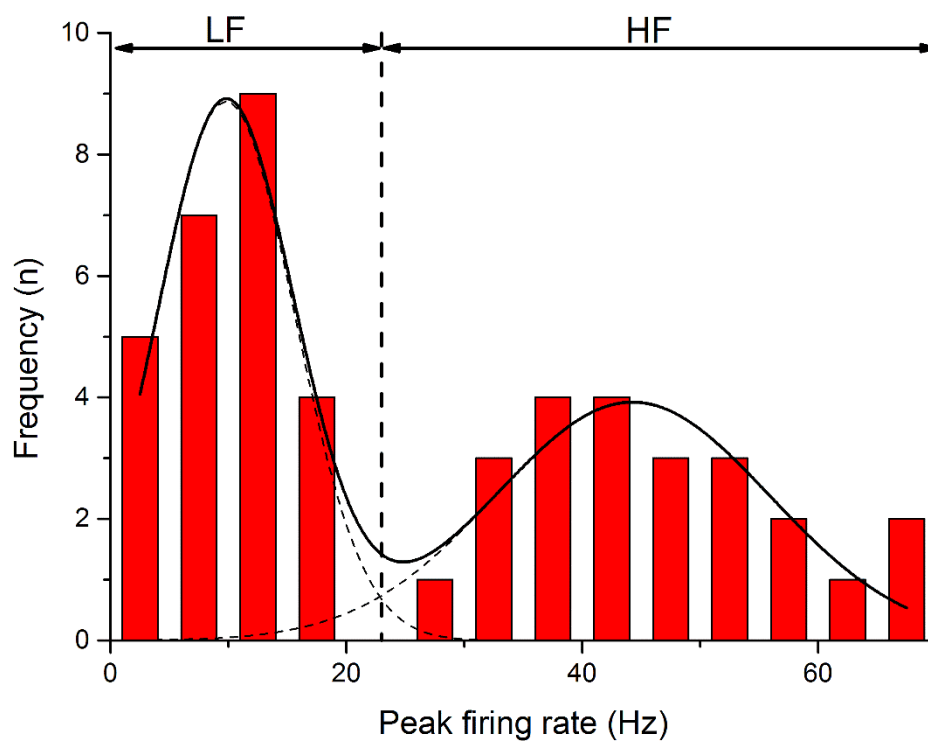
Extended Data Figure 3 | Modulation of heat responses by H_2O_2 : the role of TRPA1. **a**, Representative examples of responses to heat, in the absence and presence of H_2O_2 , in an AITC-sensitive and an AITC-insensitive sensory neuron from $DKO^{M3/V1}$ mice. **b**, Example of the suppression of heat responses in an AITC-sensitive $DKO^{M3/V1}$ neuron by the TRPA1 antagonist HC030031 ($100\mu M$). **c**, Heat-insensitive TKO sensory neurons are also insensitive to H_2O_2 ($n = 147$ from 4 mice). **d, e**, Heat responses in TKO sensory neurons (which occur in only about

2% of the total population) are neither potentiated by H_2O_2 ($400\mu M$; $n = 147$ from 4 mice) nor inhibited by 2-APB ($250\mu M$; $n = 150$ from 4 mice). Scale bars in **a–e**, 60s/100 nM. **f**, Average heat-induced increases in intracellular calcium during the three consecutive heat responses in the protocol shown in **a–d**, for the indicated subtypes of neurons. Number of tested trigeminal neurons for the data in **f**: $DKO^{M3/V1}$, $n = 308$ from 6 mice; $DKO^{M3/V1} + HC030031$, $n = 104$ from 4 mice; TKO, $n = 147$ from 4 mice. Group data are represented as mean \pm s.e.m.



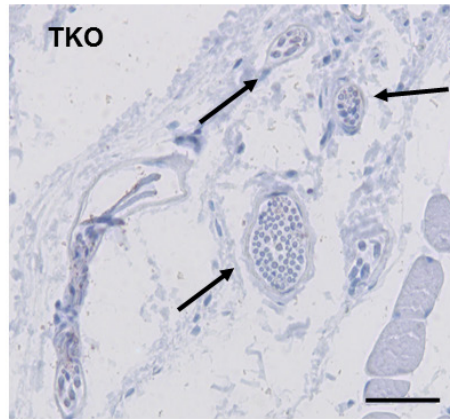
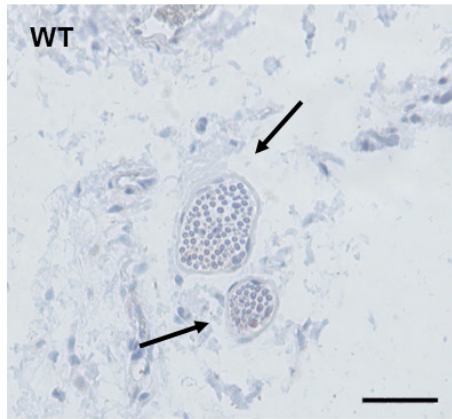
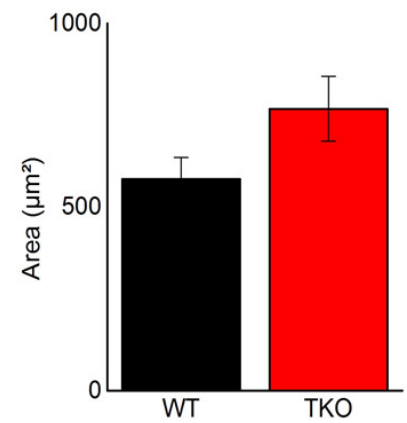
Extended Data Figure 4 | H_2O_2 potentiates heat responses in CHO cells expressing mTRPA1. **a**, Representative examples of responses to heat ($45^\circ C$), AITC ($50 \mu M$) and the ionophore ionomycin ($2 \mu M$) in CHO cells stably expressing mouse TRPA1. **b**, **c**, Potentiation of the heat response by H_2O_2 ($100 \mu M$) and inhibition by HC030031 ($100 \mu M$). **d**, Population analysis of all tested cells corresponding to the experiments in **a–c**. Control, $n = 394$; H_2O_2 , $n = 560$; $H_2O_2 + HC030031$, $n = 103$ cells.

As a conservative estimate of TRPA1-dependent heat responses, cells with a heat response larger than 5 s.d. above the mean of responses in the presence of HC030031 (dotted red line) were counted. Corresponding percentages are indicated. The H_2O_2 group was significantly different from the two other groups (** $P < 0.00001$; two-sided Mann–Whitney test with Bonferroni correction).



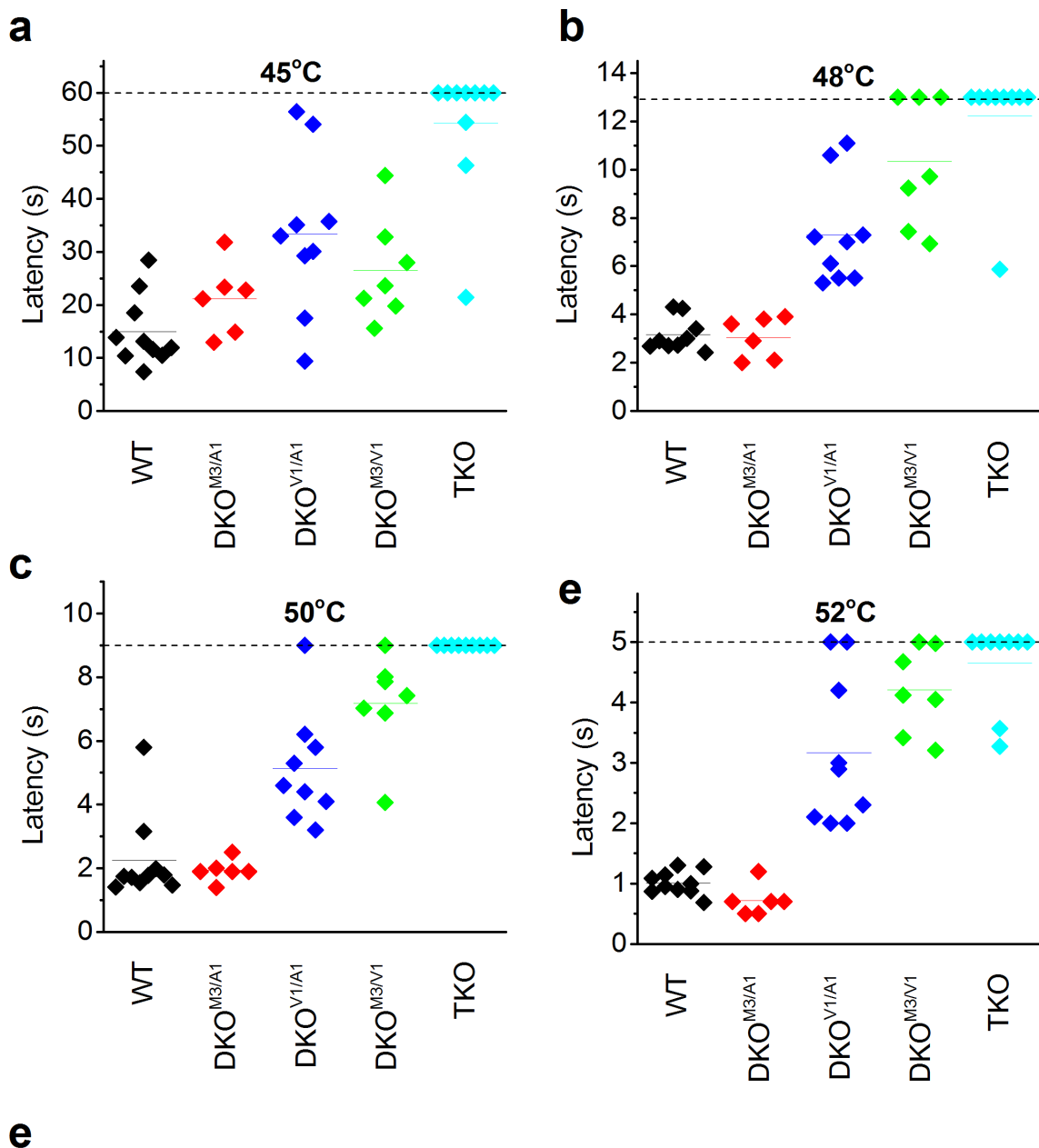
Extended Data Figure 5 | Bimodal distribution of peak heat-induced firing rates in C fibres. Histogram showing the peak firing rate of all tested heat-sensitive C fibres. Solid line shows the sum of two Gaussians

(dotted lines) fit to the data. On the basis of these data, fibres were classified as either low frequency (LF; peak firing rate < 23 Hz) or high frequency (HF; peak firing rate > 23 Hz).

a**b**

Extended Data Figure 6 | Nerve bundles in the skin. **a**, PGP9.5 staining of the plantar hind paw skin, representative of wild-type ($N=5$) and TKO ($N=5$) mice. Scale bar, 20 μm . Black arrows indicate nerve bundles. **b**, For each mouse, the nerve bundle areas of 20 sections were evaluated with

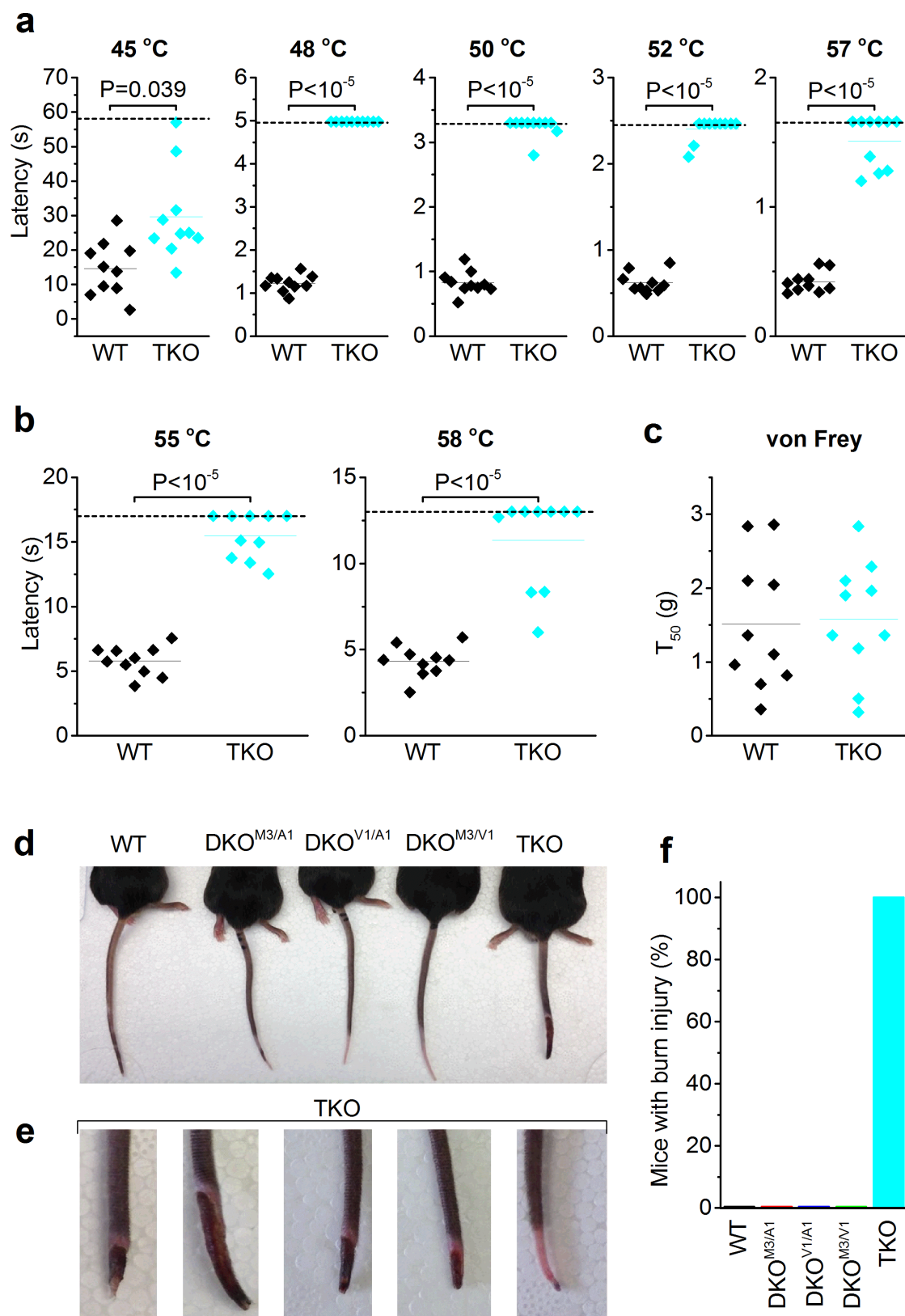
an average distance of 0.1 mm between the cross sections. No significant difference was found between the two genotypes. $P=0.08$; two-sided Student's t -test. Group data are represented as mean \pm s.e.m.



WT	P=1				
DKO ^{M3/A1}	P=0.93	P=1			
DKO ^{V1/A1}	P=0.0001	P=0.003	P=1		
DKO ^{M3/V1}	P=0.0004	P=0.006	P=1	P=1	
TKO	P<10 ⁻⁵	P<10 ⁻⁵	P<10 ⁻⁵	P<10 ⁻⁵	P=1
	WT	DKO ^{M3/A1}	DKO ^{V1/A1}	DKO ^{M3/V1}	TKO

Extended Data Figure 7 | Acute heat sensing at different temperatures. **a–d**, Withdrawal latencies for male mice of different genotypes in the tail-immersion assay at the indicated temperatures. The cut-off (dashed line) was set at four times the mean withdrawal latency of wild-type mice. Number of tested animals: wild-type, $N=10$; DKO^{M3/A1}, $N=6$; DKO^{V1/A1}, $N=9$; DKO^{M3/V1}, $N=7$; TKO, $N=9$. Horizontal lines indicate means. **e**, P values for pair-wise comparisons of the withdrawal latencies of the different genotypes. A two-way ANOVA was performed with genotype

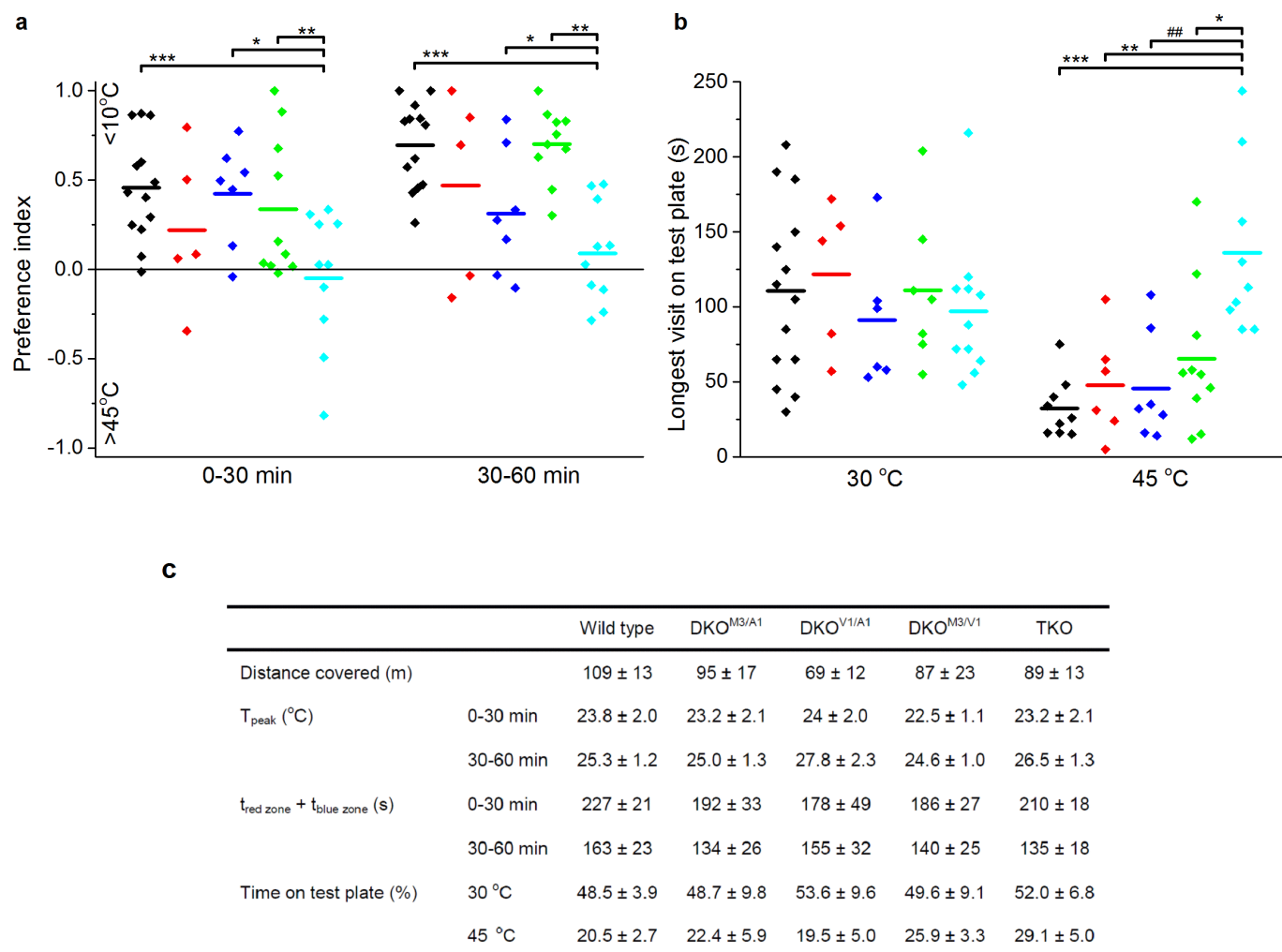
and bath temperature as factors. P values represent the result of Tukey's post-hoc tests for the factor genotype, and thus quantify global statistical differences in latency over the five tested temperatures (45, 48, 50, 52 and 57°C). Similar results, yielding $P < 10^{-4}$ for the comparison of TKO mice with either wild-type mice or mice of the different DKO genotypes, were obtained with a second independent cohort of male mice (5–6 mice per genotype).



Extended Data Figure 8 | Acute heat sensing and burn injury.

a, Withdrawal latencies of female wild-type ($N=10$) and TKO ($N=10$) mice in the tail-immersion assay at the indicated temperatures. The cut-off (dashed line) was set at four times the mean withdrawal latency of wild-type mice. Indicated P values were obtained using a two-sided Student's t -test. **b**, Withdrawal latencies of female wild-type ($N=10$) and TKO ($N=10$) mice in the hot-plate assay at the indicated temperatures. The cut-off (dashed line) was set at three times the mean withdrawal

latency of wild-type mice. Indicated P values were obtained using two-sided Student's t -tests. **c**, Sensitivity of female wild-type ($N=10$) and TKO ($N=10$) mice to calibrated von Frey hairs ($P=0.86$; two-sided Student's t -test). **d**, **e**, Photograph of the tails of male mice of the indicated genotypes (**d**) and close-up of the tails of five TKO mice (**e**) taken 3 days after a 57 °C tail-flick experiment. **f**, Percentage of mice of the indicated genotypes exhibiting visual signs of burn injury 3 days after a 57 °C tail-flick experiment (five mice per genotype).



Extended Data Figure 9 | Thermal preference. **a**, Preference index for the blue (<10°C) versus the red zone (>45°C) on the thermal gradient. Preference index was calculated as $(t_{blue\ zone} - t_{red\ zone}) / (t_{blue\ zone} + t_{red\ zone})$. *** $P = 2.2 \times 10^{-6}$, ** $P = 7.2 \times 10^{-5}$, * $P = 0.026$, two-way ANOVA with Tukey's post-hoc test for the factor genotype. Wild-type, $N = 13$; DKO^{M3/A1}, $N = 5$; DKO^{V1/A1}, $N = 7$; DKO^{M3/V1}, $N = 10$; TKO, $N = 10$. **b**, Duration

of the longest uninterrupted visit to the test plate. *** $P = 7.7 \times 10^{-5}$, ** $P = 0.0029$, # $P = 0.0013$, * $P = 0.0072$; one-way ANOVA with Tukey's post-hoc test. **c**, Parameters describing thermal preference behaviour in the thermal gradient and two-plate preference tests. Number of tested animals: wild-type, $N = 9$; DKO^{M3/A1}, $N = 6$; DKO^{V1/A1}, $N = 6$; DKO^{M3/V1}, $N = 10$; TKO, $N = 9$.

Extended Data Table 1 | General properties of TKO mice

	Wild type	TKO	P value
Body weight (g) at 12 weeks (males)	25.1 ± 0.5 (N=14)	27.0 ± 0.9 (N=13)	P=0.45
Body core temperature (°C)	37.1 ± 0.2 (N=10)	37.9 ± 0.3 (N=10)	P=0.02
Rotarod (maximal speed in rpm)	12.9 ± 1.7 (N=6)	14.4 ± 1.8 (N=6)	P=0.56
Elevated maze - time in open arms (%)	37.7 ± 3.5 (N=16)	35.9 ± 6.5 (N=16)	P=0.81
Elevated maze - distance in open arms (%)	37.2 ± 3.0 (N=16)	32.9 ± 6.4 (N=16)	P=0.55

Data are expressed as mean ± s.e.m. from *N* mice. Indicated *P* values were obtained using a two-sided Student's *t*-test.

Electromechanical vortex filaments during cardiac fibrillation

J. Christoph^{1,2,3}, M. Chebbok^{2,4}, C. Richter^{1,2,4}, J. Schröder-Schetelig^{1,2,3}, P. Bittihn⁵, S. Stein^{1,3}, I. Uzelac⁶, F. H. Fenton⁶, G. Hasenfuß^{2,4}, R. F. Gilmour Jr⁷ & S. Luther^{1,2,3,8,9,10}

The self-organized dynamics of vortex-like rotating waves, which are also known as scroll waves, are the basis of the formation of complex spatiotemporal patterns in many excitable chemical and biological systems^{1–4}. In the heart, filament-like phase singularities^{5,6} that are associated with three-dimensional scroll waves⁷ are considered to be the organizing centres of life-threatening cardiac arrhythmias^{7–13}. The mechanisms that underlie the onset, maintenance and control^{14–16} of electromechanical turbulence in the heart are inherently three-dimensional phenomena. However, it has not previously been possible to visualize the three-dimensional spatiotemporal dynamics of scroll waves inside cardiac tissues. Here we show that three-dimensional mechanical scroll waves and filament-like phase singularities can be observed deep inside the contracting heart wall using high-resolution four-dimensional ultrasound-based strain imaging. We found that mechanical phase singularities co-exist with electrical phase singularities during cardiac fibrillation. We investigated the dynamics of electrical and mechanical phase singularities by simultaneously measuring the membrane potential, intracellular calcium concentration and mechanical contractions of the heart. We show that cardiac fibrillation can be characterized using the three-dimensional spatiotemporal dynamics of mechanical phase singularities, which arise inside the fibrillating contracting ventricular wall. We demonstrate that electrical and mechanical phase singularities show complex interactions and we characterize their dynamics in terms of trajectories, topological charge and lifetime. We anticipate that our findings will provide novel perspectives for non-invasive diagnostic imaging and therapeutic applications.

Spiral waves are found ubiquitously throughout nature^{1–4}. These spiral waves form dynamically in many chemical and biological systems around two- or three-dimensional centres of rotation, which consist of singular points or filament-like lines of phase singularity^{5,6}, respectively. In the heart, spiral waves are associated with cardiac fibrillation^{7,8}, as electrical vortex-like rotating waves or rotors can be observed on the surface of the heart during arrhythmia⁹. The core regions of these rotors, which can be displayed as phase singularities using fluorescence imaging of membrane voltage or calcium^{12,13}, are thought to expand into the heart muscle where they form vortex filaments that are associated with three-dimensional (3D) electrical scroll waves⁵. However, visualization of the 3D wave phenomena that occur within the cardiac muscle has remained a major scientific challenge. Despite substantial progress in the development of tomographic optical techniques^{17–19}, the measurement of transient electrical scroll waves inside cardiac tissue has so far been impossible. Because vortex filaments are considered to be the organizing centres of fibrillation^{12,13}, insights into their dynamics

inside the heart wall are thought to be key to understanding the nature of cardiac fibrillation^{10–13} and to the conceptualization and advancement of novel techniques for the control of heart rhythm disorders^{14–16}.

Here, we present direct and fully time-resolved measurements of 3D mechanical scroll waves inside the contracting cardiac muscle of intact, Langendorff-perfused pig hearts. To elucidate the dynamics of electrical and mechanical wave phenomena in the heart, we combined panoramic fluorescence imaging (optical mapping) with high-resolution four-dimensional (4D) ultrasound imaging (Fig. 1 and Extended Data Figs 1–3). The rapid mechanical contractions within the entire ventricles during ventricular fibrillation were measured using fast 4D ultrasound imaging with a spatiotemporal resolution of 0.5–1.0 mm (field of view, approximately $10 \times 10 \times 10 \text{ cm}^3$) and 50–188 Hz volumes per second (see Methods). Tissue deformation (strain) was calculated from the 4D ultrasound data, revealing the propagation of 3D mechanical waves of contracting and dilating tissue throughout the ventricular muscle during regular rhythm and cardiac arrhythmias (Fig. 1a, d, 2a). From the amplitude of the mechanical deformation wave, the corresponding phase was computed. We induced ventricular tachycardia and ventricular fibrillation in $n = 5$ pig hearts (see Methods). During ventricular tachycardia, we observed a single stable 3D mechanical scroll wave within the left-ventricular wall (Supplementary Video 1). The phase of this 3D wave (Fig. 1a, d and Extended Data Fig. 1a, b) revealed the formation a U-shaped vortex filament, which extends from the epicardial surface to deep inside the left ventricular wall and back. On the surface of the heart, the filament ended in two phase singularities with opposite topological charges, corresponding to two stable counter-rotating mechanical spiral waves. During ventricular fibrillation, drifting U-shaped mechanical filaments were observed during ventricular fibrillation (Fig. 2a). In this example, the breakthrough of the filament through the epicardial surface resulted in the formation of two transient I-shaped filaments and two corresponding counter-rotating, outwardly drifting mechanical scroll waves (Supplementary Video 2 and Extended Data Fig. 1c). The formation, transition and extinction of U-shaped, I-shaped and O-shaped mechanical filaments was also observed during ventricular fibrillation (Fig. 2b), phenomena similar to those predicted by models for electrical filaments⁵. The mechanical wave phenomena that were obtained from tissue deformation analyses using ultrasound imaging were compared to the electrical activation and mechanical motion data that were obtained from simultaneous multimodal optical mapping of the surface of the heart (see Supplementary Video 9 and Extended Data Figs 4d, e, 9). The optical mapping data confirmed the co-existence of rotating waves of mechanical deformation and electrical activation, which appeared simultaneously within the same region of the heart and exhibited the same rotational sense and rotation period, and retained a similar morphological shape

¹Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany. ²German Center for Cardiovascular Research (DZHK), Partner Site Göttingen, Göttingen, Germany. ³Institute for Nonlinear Dynamics, University of Göttingen, Göttingen, Germany. ⁴Department for Cardiology and Pneumology, University Medical Center Göttingen, Göttingen, Germany. ⁵BioCircuits Institute, University of California San Diego, La Jolla, California, USA. ⁶School of Physics, Georgia Institute of Technology, Atlanta, Georgia, USA. ⁷University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada. ⁸Institute of Pharmacology, University Medical Center Göttingen, Göttingen, Germany. ⁹Department of Bioengineering, Northeastern University, Boston, Massachusetts, USA. ¹⁰Department of Physics, Northeastern University, Boston, Massachusetts, USA.

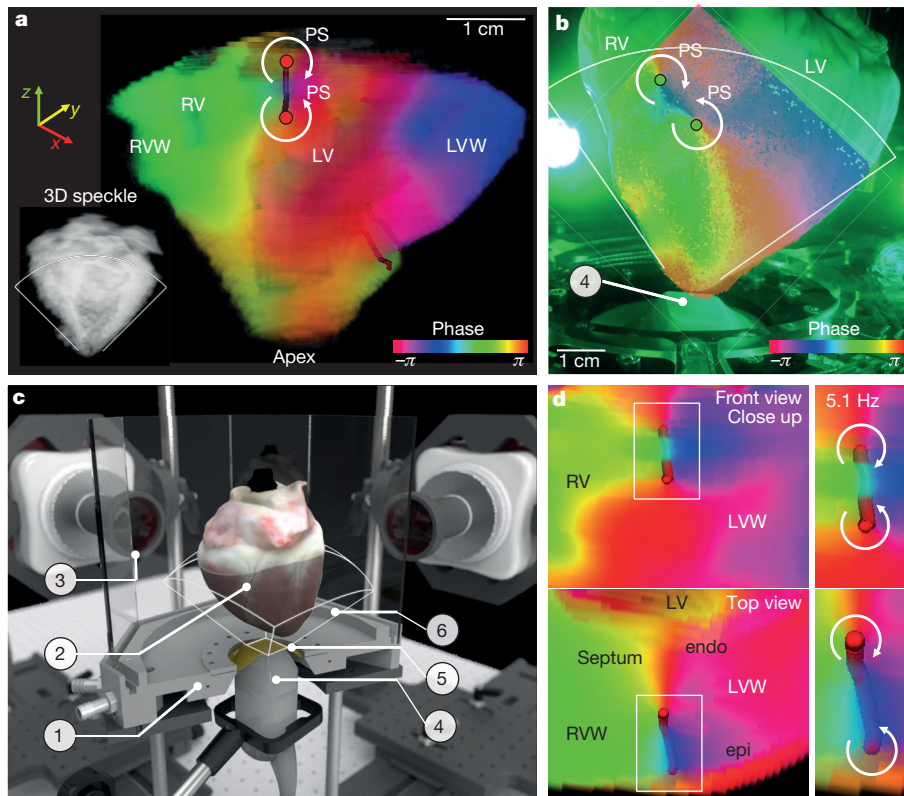


Figure 1 | Measurement of intramural mechanical filament-like phase singularity. **a**, Double scroll wave (phase) and U-shaped mechanical filament (red) recorded at 134 volumes per second during ventricular tachycardia inside the left ventricular wall (LVW) of a contracting Langendorff-perfused pig heart using 4D ultrasound (Supplementary Video 1 and Extended Data Fig. 1 a, b). LV, left ventricle; RV, right ventricle; RVW, right ventricular wall; PS, phase singularity. **b**, Figure-of-eight spiral wave (membrane voltage) on the surface of the left and right ventricle with two phase singularities obtained by recording

optical mapping data at the same time as the recording shown in **a**. The ultrasound transducer is indicated by (4); the white lines indicate the field of view. **c**, Schematic of the experimental setup. (1) temperature-controlled perfusion bath; (2) Langendorff-perfused pig heart; (3) high-speed EMCCD (electron-multiplying charge-coupled device) cameras (only two shown) and LED illumination (not shown); (4) 4D ultrasound transducer; (5) acoustic window (membrane); (6) 3D field of view of ultrasound transducer. See the Methods for details. **d**, Detail (front and top view) of the U-shaped mechanical filament (red) of the scroll wave shown in **a**.

(Fig. 2a and Extended Data Fig. 9c). In general, we found that the cardiac muscle exhibits deformations during cardiac fibrillation²⁰ that retain vortex-like patterns and, like fingerprints of electrical vortex activity, reveal the topological organization of cardiac fibrillation throughout the heart wall. Furthermore, we found that rotational centres of the mechanical rotor patterns can be displayed as phase singular points or filament-like lines of phase singularity, as has been shown previously for electrical activity^{12,13} (see Supplementary Videos 1, 2, 6–8, 11).

To further elucidate the dynamics of electrical and mechanical vortex filaments, we conducted computer simulations of scroll waves in excitable and contractile cardiac tissue (see Methods). Figure 2c, d show simulation data of electromechanical vortex filament dynamics using realistic heart geometry and fibre orientation. Our simulations indicate that electromechanical turbulence in the heart is composed of entangled pairs of co-existing electromechanical vortex filaments. The filaments emerge in close vicinity to each other due to electromechanical coupling (Supplementary Video 3). The simulations suggest that the experimentally observed mechanical filaments inside the ventricular wall may be intrinsic dynamical entities that can be used to characterize the spatiotemporal organization of cardiac fibrillation.

To investigate the dynamics and interaction of electrical and mechanical rotors experimentally, we have developed a multi-modal fluorescence imaging system, which allows for the simultaneous measurement of membrane voltage, intracellular calcium and mechanical strain on the surface of the heart (Fig. 3b–d). Standard optical mapping is very sensitive to mechanical motion of tissue, which may result in

substantial distortions of the signal. To avoid these motion artefacts²¹, pharmacological excitation–contraction uncoupling agents have been applied in previous experiments to inhibit any contractile motion of the heart. By contrast, we used computer vision techniques to disentangle fluorescence signals (voltage, calcium) from mechanical deformation, which not only overcomes the limitations imposed by motion artefacts^{22–26}, but also enables quantitative motion analysis (Supplementary Videos 4, 5). Figure 3b illustrates the marker-free motion-tracking algorithm. The algorithm estimates the displacement of the epicardial surface from a reference configuration to its instantaneous configuration at any subsequent time t . This transformation of coordinates is used to convert the fluorescence signal from the stationary camera frame of reference into the co-moving frame of reference. Within the co-moving frame of reference, fluorescence signals are essentially artefact-free. The efficacy and quality of the motion tracking may be assessed from the examples shown in Fig. 3c. The original fluorescence data show that the signal is substantially distorted by tissue motion, rendering a quantitative analysis impossible. However, within the co-moving reference frame, artefacts are significantly reduced, allowing the voltage and calcium signals to be reliably retrieved. Simultaneously, non-rigid motion tracking provides a measurement of the tissue coordinates and local deformation of the cardiac tissue as a function of time. An example of a simultaneous measurement of membrane voltage, calcium and the corresponding strain rate is shown in Fig. 3d (Supplementary Video 8). The time series show that the tissue experiences deformations that correlate with the sequence of electrical activation.

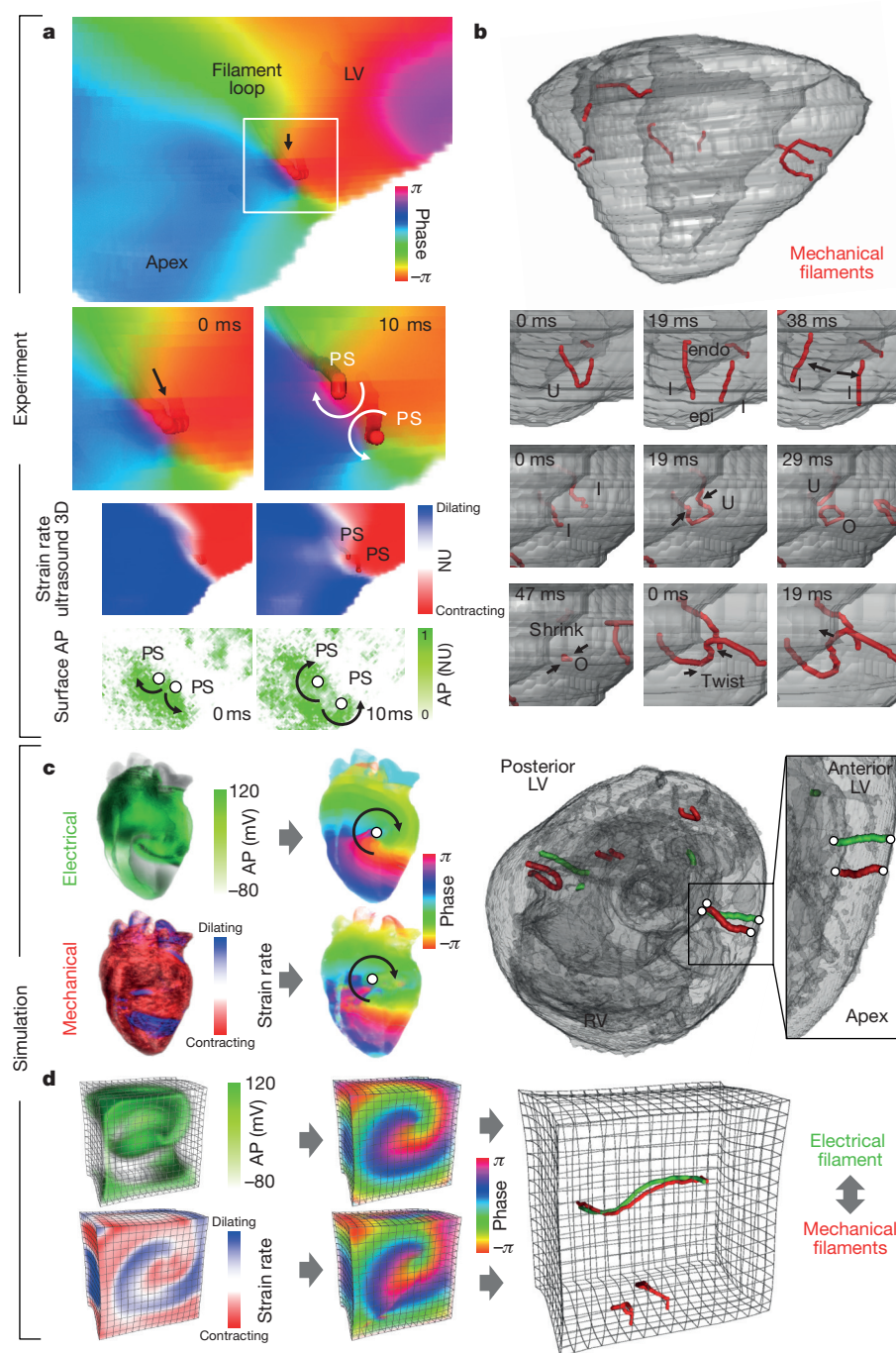


Figure 2 | Electromechanical phase singularity in the experiments and simulations. **a**, Breakthrough of the U-shaped intramural mechanical filament (red) on the epicardial surface (LVW, pig) results in the formation of two transmural filaments during ventricular fibrillation and two phase singularities (circles) with opposite charges (counter-rotating spirals) on the surface. Optical mapping (voltage and strain rate) confirms the breakthrough of filament that is observed with 4D ultrasound (Supplementary Video 2). **b**, Creation and extinction of mechanical filaments (red) inside pig ventricles (grey shaded) between endocardial

(endo) and epicardial (epi) heart surface during ventricular fibrillation imaged using 4D ultrasound. Snapshots show interaction and transition of I-, U-, and O-shaped filaments. **c**, **d**, Numerical simulation of electrical (green, action potential (AP)) and mechanical (red-blue: strain rate with contracting (red) and dilating (blue) rates of deformation) scroll waves in the rabbit heart (computed tomography scan) and rectangular bulk with associated electrical (green) and mechanical (red) vortex filaments (Supplementary Video 3). NU, normalized units.

Applying this optical mapping system to ventricular fibrillation, we observed simultaneously co-localized rotors in the action potential, calcium and strain measurements (Fig. 3e, f) on the surface of intact Langendorff-perfused rabbit hearts (Supplementary Videos 6–8). The coupling of cardiac excitation and contraction (Fig. 3a) may result in a mechanical rotor, which consists of two adjacent regions of dilating and contracting strain rates rotating around each other (see

also Extended Data Fig. 4a–c). Analogous to electrical rotors, mechanical rotors are characterized by a phase singularity, indicating the centre of rotation. In order to investigate the spatiotemporal organization of electrical and mechanical rotors, data and statistical analyses were obtained in rabbit ($n = 3$) and pig hearts ($n = 3$). During ventricular fibrillation, the position of electrical and mechanical rotors were calculated and tracked over time. In total, 3,180 trajectories of phase

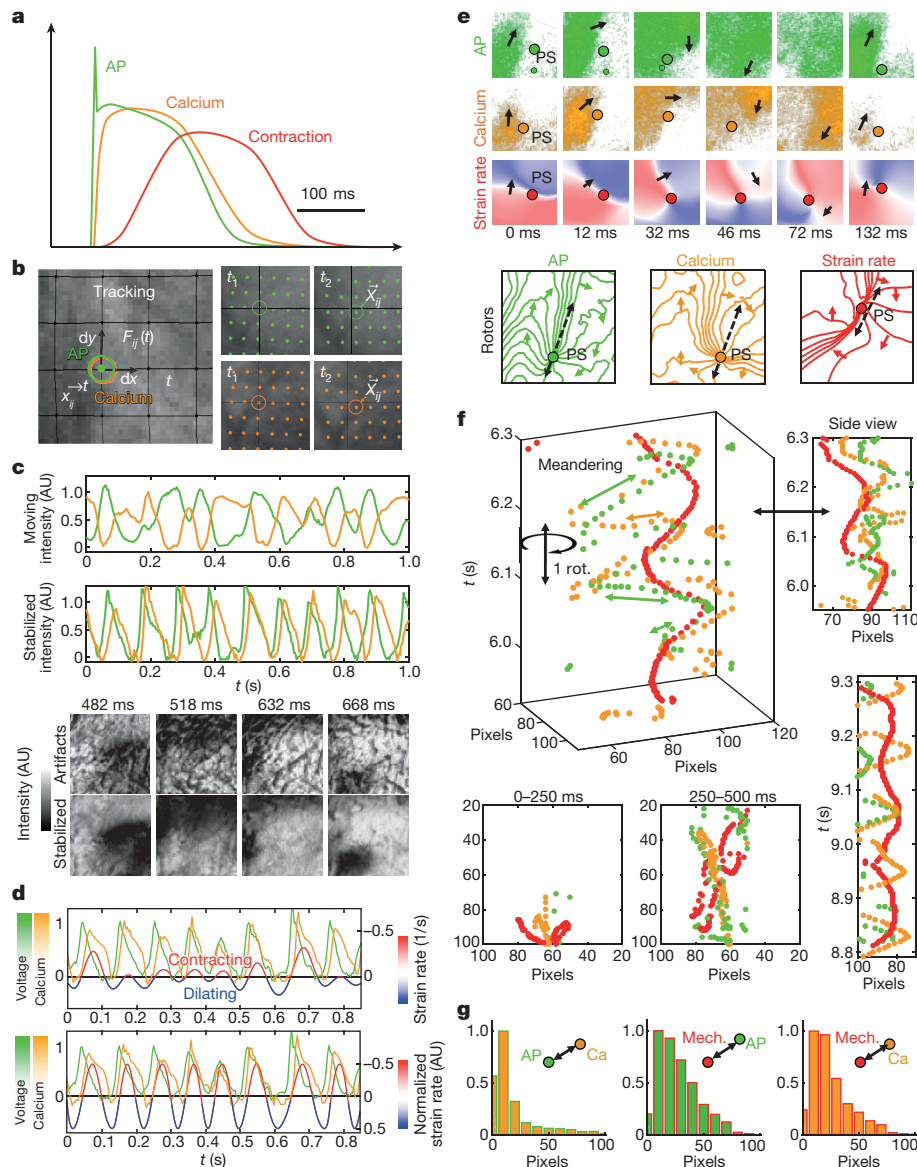


Figure 3 | Optical mapping of electromechanical phase singularities during ventricular fibrillation. **a**, Excitation–contraction coupling²⁸ in the cardiomyocyte. Action potentials (AP, green) trigger the release of intracellular calcium (orange), resulting in cellular contraction (red). **b**, Numerical motion tracking in optical mapping data disentangles mechanical motion (deformation) from voltage and calcium signals. **c**, Efficacy of numerical motion tracking and motion-artefact removal. Top graph, without motion tracking, the time-series shows artefacts. Bottom graph, action potentials and calcium measured on the contracting heart surface after motion tracking. Top image sequence, the non-tracked action potential wave shows substantial motion artefacts (with a high frequency spatial pattern). Bottom image sequence, after motion tracking, a spiral wave (action potential, dark indicates depolarized tissue) is revealed (Supplementary Videos 4, 5 and Extended Data Fig. 4a). **d**, Simultaneous measurement of action potentials, calcium and deformation during ventricular fibrillation using optical mapping.

singularities (electrical and mechanical) were measured in pig hearts, and 1,142 trajectories of phase singularities were measured in rabbit hearts. Along these trajectories, the positions of the phase singularities were sampled at 250 or 500 Hz. Representative trajectories of the electrical and mechanical phase singularities (x – y – t plots) are shown in Figs 3f, 4f (rabbit) and Fig. 4a–c (pig). From these trajectories, basic statistical information has been retrieved (Figs 3g, 4d and Extended Data Fig. 6a–d).

Repetitive electrical activation leads to a corresponding sequence of dilating (blue indicates strain rates <0) and contracting (red indicates strain rates >0) rates of deformation. Top, raw unnormalized strain rate signal containing amplitude variations. Bottom, normalized strain rate used for computing of the phase signal (Extended Data Fig. 4b). **e**, Rotating electrical (voltage and calcium) and mechanical activity during ventricular fibrillation. **f**, Phase singularity space–time plots of action potentials, calcium and mechanics on the ventricular surface. Mechanical phase singularities co-localize with voltage and calcium phase singularities over multiple rotations, even for meandering rotors (1 pixel (px) = 0.16 mm, 100 pixels = approximately the diameter of the heart (16 mm)). **g**, Histogram of distances between phase singularities (voltage-to-calcium, voltage-to-mechanics, calcium-to-mechanics) during ventricular fibrillation shows co-localization with peaks at approximately 0.8 ± 0.8 mm (average distance between phase singularities is 4.0 ± 0.8 mm). AU, arbitrary units.

Our experimental data show that across hearts and species (pig, rabbit) the observed electrical and mechanical phase singularities remain similar in number and lifetime (Fig. 4g, h). The data also support the hypothesis that, on average, electrical phase singularities are associated with nearby mechanical phase singularities, with the average nearest neighbour distance between an electrical and its corresponding mechanical phase singularity being small, compared to the distance to the next electrical or mechanical phase singularity (Fig. 4a–e and Extended Data Table 1).

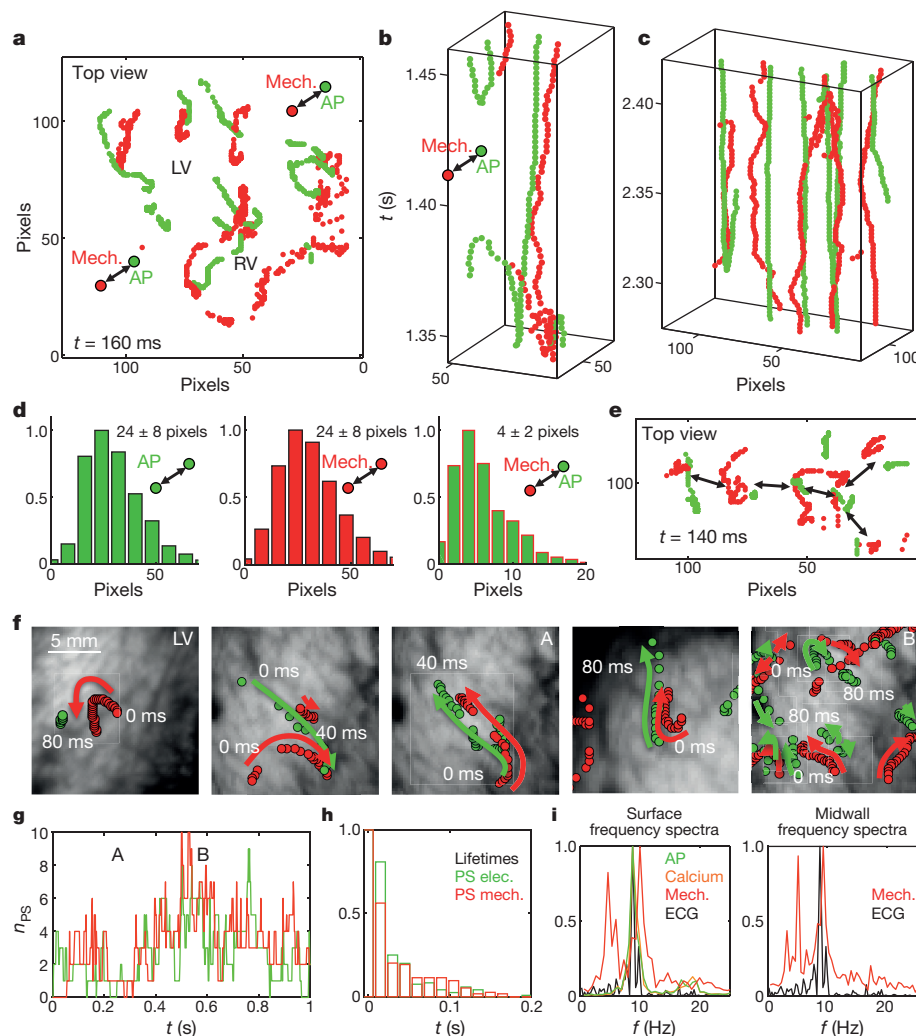


Figure 4 | Dynamics of electromechanical phase singularity during ventricular fibrillation. **a–c, e,** Co-localization of epicardial voltage (green) and mechanical (red) phase singularities on the pig heart surface at 160-ms (**a**), 110-ms (**b**), 120-ms (**c**) and 140-ms (**e**) intervals (1 pixel = 0.78 mm). **d,** Histograms of nearest neighbour distances between phase singularities (voltage-to-voltage, mechanics-to-mechanics and voltage-to-mechanics) during ventricular fibrillation (pig). **f,** Phase singularity trajectories within intervals of 40 ms and 80 ms during ventricular fibrillation (rabbit, $n = 3$). **g,** Temporal evolution of electrical

and mechanical phase singularity numbers (rabbit). **f, g,** A and B indicate intervals of low and high phase singularity density, respectively.

h, Histogram of lifetimes of electrical and mechanical phase singularities (rabbit). **i,** Frequency spectra of voltage, calcium and strain rate on the rabbit heart surface obtained from optical mapping, mechanical deformation (strain rate) within the heart wall obtained from 2D ultrasound and electrocardiogram recordings. Dominant frequencies are 9 ± 1 Hz (Extended Data Figs 4e, 9a–c).

On the cellular level, electrical and mechanical processes in the heart are related through excitation–contraction coupling and mechanoelectrical feedback. On the organ level, diffusive coupling of cardiomyocytes allows electrical impulses to propagate as an activation wave, which leads to well-coordinated and relatively uniform mechanical contraction of cardiac tissue in the normal heart. In the diseased heart, excitation–contraction coupling and mechanoelectrical feedback may be altered, resulting in contractile dysfunction and onset of arrhythmias. Although electrical and mechanical processes are intrinsically coupled, the physical nature of the emergent electrical and mechanical wave phenomena is very different. Electrical waves in the heart are solitary excitation waves (and, for example, annihilate under collision) with a typical conduction velocity of 1 m s^{-1} . By contrast, electromechanical waves²⁷ are elastic deformation waves, which are initiated by the excitation waves via excitation–contraction coupling. However, their wave-like propagation may be altered owing to subsequent interactions between the mechanically interconnected physiological structures of the heart. Passive elastic waves travelling through the heart also have the potential to effectively introduce non-local mechanical interactions because of the large speed of sound

(approximately $1,500 \text{ m s}^{-1}$) in cardiac tissue. Although our results for ventricular fibrillation and ventricular tachycardia provide evidence for coupling of the voltage and calcium dynamics (Extended Data Fig. 5), coupled electromechanical dynamics and the coexistence of electrical and mechanical phase singularities, elastic patterns should not be considered as a simple proxy for the underlying electrical activation. Consequently, we expect that the mutual coupling of electrical and mechanical waves in a deformable excitable medium, such as the heart, gives rise to very rich spatiotemporal dynamics that still need to be explored (Supplementary Video 12, Extended Data Figs 7, 8 and Supplementary Discussion). Despite their complex interactions, electrical and mechanical phase singularities show similar properties, including topological charge, structure, dynamics and lifetime, suggesting that the spatiotemporal dynamics of the electrical and mechanical system are essentially two sides of the same coin.

We believe that the imaging approach introduced in this study will provide the basis for the integrative study of these electromechanical phenomena (such as excitation–contraction coupling and mechanoelectrical feedback) at the tissue and organ level in health and disease. We expect that these developments will help to unravel the mechanisms

that underlie the spatiotemporal organization of electromechanical turbulence during cardiac fibrillation and could lead to safer and more effective diagnostic and therapeutic clinical applications.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 December 2016; accepted 14 February 2018.

Published online 21 February 2018.

1. Welsh, B. J., Gomata, J. & Burgess, A. E. Three-dimensional chemical waves in the Belousov–Zhabotinskii reaction. *Nature* **304**, 611–614 (1983).
2. Rotermund, H. H., Engel, W., Kordes, M. & Ertl, G. Imaging of spatio-temporal pattern evolution during carbon monoxide oxidation on platinum. *Nature* **343**, 355–357 (1990).
3. Huang, X. *et al.* Spiral waves in disinhibited mammalian neocortex. *J. Neurosci.* **24**, 9897–9902 (2004).
4. Winfree, A. T. Scroll-shaped waves of chemical activity in three dimensions. *Science* **181**, 937–939 (1973).
5. Fenton, F. & Karma, A. Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: filament instability and fibrillation. *Chaos* **8**, 20–47 (1998).
6. Clayton, R. H., Zhuchkova, E. A. & Panfilov, A. V. Phase singularities and filaments: simplifying complexity in computational models of ventricular fibrillation. *Prog. Biophys. Mol. Biol.* **90**, 378–398 (2006).
7. Winfree, A. T. Electrical turbulence in three-dimensional heart muscle. *Science* **266**, 1003–1006 (1994).
8. Krinsky, V. I. Spread of excitation in an inhomogeneous medium (state similar to cardiac fibrillation). *Biophysika* **11**, 676–683 (1966).
9. Davidenko, J. M., Pertsov, A. V., Salomonsz, R., Baxter, W. & Jalife, J. Stationary and drifting spiral waves of excitation in isolated cardiac muscle. *Nature* **355**, 349–351 (1992).
10. Gray, R. A. *et al.* Mechanisms of cardiac fibrillation. *Science* **270**, 1222–1223 (1995).
11. Jalife, J. & Gray, R. Drifting vortices of electrical waves underlie ventricular fibrillation in the rabbit heart. *Acta Physiol. Scand.* **157**, 123–132 (1996).
12. Gray, R. A., Pertsov, A. M. & Jalife, J. Spatial and temporal organization during cardiac fibrillation. *Nature* **392**, 75–78 (1998).
13. Witkowski, F. X. *et al.* Spatiotemporal evolution of ventricular fibrillation. *Nature* **392**, 78–82 (1998).
14. Fenton, F. H. *et al.* Termination of atrial fibrillation using pulsed low-energy far-field stimulation. *Circulation* **120**, 467–476 (2009).
15. Luther, S. *et al.* Low-energy control of electrical turbulence in the heart. *Nature* **475**, 235–239 (2011).
16. Janardhan, A. H. *et al.* A novel low-energy electrotherapy that terminates ventricular tachycardia with lower energy than a biphasic shock when antitachycardia pacing fails. *J. Am. Coll. Cardiol.* **60**, 2393–2398 (2012).
17. Mitrea, B. G., Caldwell, B. J. & Pertsov, A. M. Imaging electrical excitation inside the myocardial wall. *Biomed. Opt. Express* **2**, 620–633 (2011).
18. Mitrea, B. G., Wellner, M. & Pertsov, A. M. Monitoring intramyocardial reentry using alternating transillumination. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2009**, 4194–4197 (2009).
19. Baxter, W. T., Mironov, S. F., Zaitsev, A. V., Jalife, J. & Pertsov, A. M. Visualizing excitation waves inside cardiac muscle using transillumination. *Biophys. J.* **80**, 516–530 (2001).
20. Wiggers, C. J. The mechanism and nature of ventricular fibrillation. *Am. Heart J.* **20**, 399–412 (1940).
21. Svrcek, M., Rutherford, S., Chen, A. Y. H., Provaznik, I. & Smaill, B. Characteristics of motion artifacts in cardiac optical mapping studies. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2009**, 3240–3243 (2009).
22. Rohde, G. K., Dawant, B. M. & Lin, S. F. Correction of motion artifact in cardiac optical mapping using image registration. *IEEE Trans. Biomed. Eng.* **52**, 338–341 (2005).
23. Seo, K. *et al.* Structural heterogeneity in the ventricular wall plays a significant role in the initiation of stretch-induced arrhythmias in perfused rabbit right ventricular tissues and whole heart preparations. *Circ. Res.* **106**, 176–184 (2010).
24. Bourgeois, E. B., Bachtel, A. D., Rogers, J. M., Huang, J. & Walcott, G. P. Simultaneous optical mapping of transmembrane potential and wall motion in isolated, perfused whole hearts. *J. Biomed. Opt.* **16**, 096020 (2011).
25. Zhang, H., Iijima, K., Huang, J., Walcott, G. P. & Rogers, J. M. Optical mapping of membrane potential and epicardial deformation in beating hearts. *Biophys. J.* **111**, 438–451 (2016).
26. Christoph, J., Schröder-Schotelig, J. & Luther, S. Electromechanical optical mapping. *Prog. Biophys. Mol. Biol.* **130**, 150–169 (2017).
27. Provost, J., Lee, W.-N., Fujikura, K. & Konofagou, E. E. Imaging the electromechanical activity of the heart in vivo. *Proc. Natl Acad. Sci. USA* **108**, 8565–8570 (2011).
28. Bers, D. M. Cardiac excitation–contraction coupling. *Nature* **415**, 198–205 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank U. Parlitz and N. Otani for discussions, M. Kunze, D. Hornung and U. Schminke for technical assistance, L. Bess for linguistic advice and FujiFilm VisualSonics Inc. and Siemens HealthCare for their technical support. This work was supported by the German Ministry for Education and Research through FKZ 031A147 (S.L.); by the German Center for Cardiovascular Research (S.L. and G.H.); by the German Research Foundation through SFB 1002 Modulatory Units in Heart Failure (S.L., J.C. and G.H.) and SFB 937 Collective Behavior of Soft and Biological Matter (S.L. and J.C.); by the National Science Foundation (NSF) grants 1341190 and 1413037 (F.H.F.); by the American Heart Association (AHA) grant 15POST25700285 (I.U.); by the Human Frontiers Science Program through fellowship LT000840/2014-C (P.B.) and by the Max Planck Society.

Author Contributions J.C. and S.L. designed research and wrote the paper with F.H.F., G.H., P.B. and R.F.G.; J.C., S.L. and J.S.-S. designed the experiments; I.U. and F.H.F. contributed to the experimental setup; J.C. performed the experiments and analysed the data; M.C., J.S.-S., C.R., I.U. and F.H.F. contributed to the experiments; J.C., P.B. and S.S. provided the numerical simulations. All authors read and approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.C. (jan.christoph@ds.mpg.de) and S.L. (stefan.luther@ds.mpg.de).

Reviewer Information *Nature* thanks A. Holden, J. Jalife, E. Konofagou and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Animal ethics. All procedures regarding care and use of animals were solely performed by appropriately trained personnel and were in accordance with German animal welfare laws and were reported to the local animal welfare representatives. Pre-experimental handling and humane welfare-oriented procedures strictly followed the regulations for animal welfare, in agreement with German legislation, local stipulations and in accordance with recommendations of the Federation of European Laboratory Animal Science Associations (FELASA). All involved scientists and technicians have been accredited by the responsible ethics committee (the Lower Saxony State Office for Customer Protection and Food Safety—LAVES).

Ex vivo electromechanical imaging. Experiments were conducted in isolated, intact Langendorff-perfused pig ($n = 5$) and rabbit ($n = 5$) hearts. Ventricular fibrillation and ventricular tachycardia were induced using rapid pacing^{14,15}. Hearts were simultaneously imaged at high speeds using optical fluorescence (optical mapping) and ultrasound imaging.

Optical fluorescence imaging. Fluorescence imaging was performed using either a panoramic imaging setup with 4 high-speed cameras (pig, together with 4D ultrasound) imaging voltage only (mono-parametric) or a monocular single-camera setup (rabbit, together with 2D ultrasound) filming voltage and calcium simultaneously (multi-parametric), as previously described²⁹. Imaging was performed at a frame rate of 500 Hz with a spatial resolution of 0.16–0.78 mm with 2–10-cm fields of view with high-speed EMCCD cameras (Evolve, 128 × 128 pixels, Photometrics) together with high-numerical-aperture lenses (pig: 9 mm, f1.4, Fujinon; rabbit: 52 mm, f0.95–16, Tokina). Hearts were imaged from the side through the glass walls of an eight-sided aquarium at a distance of about 20.0–40.0 cm. During mono-parametric imaging of voltage, green excitation light was provided by up to 12 high-performance light-emitting diodes (LEDs, 532 nm ± 10 nm centre wave length). During multi-parametric imaging of voltage and calcium, two sets of high-performance LEDs (three green and three red LEDs, LZ4-G40-100 and LZ4-R40-100, LED Engin; filtered at 550 ± 5 nm and 640 ± 5 nm) were switched rapidly (500 Hz) using a custom-made fast-switching diode driver box (I.U. and F.H.F., Georgia Institute of Technology), resulting in an effective frame rate of 250 Hz for voltage and calcium recordings, respectively. Emission light was filtered using red/near-infrared longpass or bandpass filters (mono-parametric imaging, longpass filter with transmission > 610 nm; multi-parametric imaging, dual-bandpass filter with transmission bandwidths at 585 nm and 800 nm, ET585/50 nm and 800/200 nm; Chroma Technologies). In mono-parametric imaging mode, the tissue was stained with a voltage-sensitive dye (Di-4-ANEPPS, 20 ml of 1 mMol l⁻¹ concentrated dye in Tyrode solution). In multi-parametric imaging mode, the tissue was stained with a voltage-sensitive dye (Di-4-ANBDQPC, 40 µl bolus injection of 24.4 mg ml⁻¹ concentrated dye in ethanol) and a calcium-sensitive dye (Rhod2-AM, 500 µl and 4.5 ml Tyrode solution over 5 min injected into the bubble trap). Pharmacological excitation–contraction uncoupling substances were not used. Cameras were triggered externally (20 MHz Function/Waveform Generator, 33220A, Agilent) and trigger signals were captured using a data-acquisition system (MP150/AcqKnowledge, Biopac Systems Inc.). Movies were acquired using custom-made recording software (MultiRecorder, version 2.0.9, Research Group Biomedical Physics, Max Planck Institute for Dynamics and Self-Organization).

Motion tracking, motion compensation and strain analysis in fluorescence video data. The heart's motion was tracked using Lucas–Kanade optical flow estimation algorithms^{30–32}. The contrast in the video data was enhanced to facilitate motion tracking, as previously described²⁶. The tissue's mechanical configuration was tracked in every pixel throughout the sequence of images with respect to one selected reference frame. The resulting time-varying, two-dimensional displacement vector fields were stored for further processing and erroneously tracked displacements were determined by computing the deviations of the vectors in a small neighbourhood of pixels (3–5 pixels diameter) and outliers were removed based on a threshold value ($tr = 5.0$ accumulated length in pixels) or by removing individual vectors with lengths that exceeded a threshold value ($tr = 4.0$ pixels). The missing vectors (less than 5%) were substituted by a vector interpolated from the remaining surrounding displacement vectors (within 5–7 pixels diameter). The resulting time-varying vector field was smoothed in space and over time using spatiotemporal filter kernels with a Gaussian profile (diameter $d = 3–5$ pixels, 5–7 frames linear averaging). To obtain the instantaneous displacements between two subsequent frames, each frame was subtracted from the previous frame. The resulting time-varying vector field of instantaneous displacements was smoothed in space and over time using spatiotemporal filter kernels with a Gaussian profile (radius $r = 20–30$ pixels, 5–11 frames linear averaging). From the smoothed vector field the gradient deformation tensors and the Green–Lagrangian strain tensors were computed. The data of the tensors were reduced to scalar-valued

data calculating tensor invariants, that is, the tensor trace, principal stretches or eigenvalues. The strain invariant (tensor trace) was normalized using sliding-window normalization as shown in Fig. 3d or Extended Data Fig. 4b, with a window width w of at least $1 \times$ the period of the dominant frequency and typically not more than 1.5–2.5 periods of the dominant frequency. The resulting time-varying strain rate map was smoothed in space and over time using spatiotemporal filter kernels with a Gaussian profile ($d = 5$ pixels, 7 frames linear averaging). Warped, motion-stabilized videos were created using the evolution of the tracked tissue configurations to redistribute the intensity distribution of each pixel from the current frame along the inversely shifted direction to the tissue configuration of the reference frame. Pixel values were resampled into a regular pixel grid corresponding to the reference configuration using polygon clipping. All computational routines for extraction, processing and visualization of the fluorescence imaging data were custom-made routines written in C/C++ and OpenGL.

Ultrasound imaging. Ultrasound imaging was performed either with a 4D ultrasound probe (4Z1c, matrix-array transducer, 2.8 Mhz centre frequency) and system (Acuson sc2000, Siemens AG) in pig hearts, or a 2D cross-sectional probe (MS-550D, linear array transducer, 22–55 MHz broadband frequency) and system (Vevo 2100, FUJIFILM Visualsonics Inc.) in rabbit hearts. 4D imaging refers to the acquisition of 3D volume images over time (3D + t). 2D imaging refers to the acquisition of 2D images over time (2D + t , B-mode). The 4D probe produces volume frames with a pyramidal shape (maximum opening angle 90 × 90 degrees, 6–14 cm lateral depth or pyramid height) and a relatively coarse spatial resolution (approximately 100 voxels³) but large fields of views (approximately 3 × 6 × 6 cm³ to 10 × 10 × 6 cm³) at volume frame rates ranging from 51 volumes per second to 188 volumes per second depending on the lateral depth and opening angle of the pyramid (see Supplementary Table 1 for details). The spatial resolution is approximately 0.5 mm³ and decreases with the lateral distance from the transducer. The probe was positioned below an acoustic window (latex membrane) centred at the bottom of the tissue bath (Fig. 1c). A transducer mounting system (custom-made mounting arm based on dial indicator holder arm, Hoffmann Group) allowed arbitrary alignment of the imaging field of view (around several axes) and to secure the transducer in a stationary position once the desired imaging position and alignment was found. The field of view of the ultrasound imaging was facing upwards, covering the ventricles of the pig heart (Fig. 1b, c). The entire ventricular walls were imaged at a temporal resolution of 5.3–11.8 ms (6–10 cm lateral field of view, 85–188 volumes per second). With a reduced field of view (that is, 36° × 90° to 50° × 90°), the entire left ventricular wall was imaged at a temporal resolution of 5.3–7.5 ms (134–188 volumes per second). The sampling distance during high-speed 4D imaging was reduced to enhance imaging speeds (T2 pre-setting from manufacturer). The 2D probe produces high-resolution (13 µm in lateral and 27 µm in transversal direction) cross-sectional (B-mode) ultrasound speckle images. Imaging was performed at low line densities (pre-setting: 128 lines) at frame rates ranging from 279 to 309 frames per second with fields of views ranging in the order of approximately 1 cm × 1 cm and the frame rate varying with B-mode image widths and depths (see Supplementary Table 1 for details). The probe was inserted into the bath from the top facing the bottom of the aquarium. Hearts were imaged with the 2D imaging plane embedded within the left ventricular wall, intersecting the wall tangentially approximately at midwall. A transducer mounting system allowed arbitrary alignment of the field of view (around several axes) and fixing the transducer in a stationary position. The imaging plane normally approximately coincided with the optical axis of the fluorescence imaging setup. Ultrasound TTL-trigger signals were captured using a data-acquisition system and software (MP150/AcqKnowledge, Biopac Systems Inc.), to be able to synchronize the start of the recording with other acquired data.

Ultrasound data post-processing. 2D and 3D ultrasound movies were acquired in RF-format (proprietary formats from manufacturers, FUJIFILM Visualsonics Inc. and Siemens AG) and saved as 3D and 4D arrays (custom binary format, 8-bit brightness mode, speckle intensities as integer values ranging from 0 to 255). The arrays were then converted to 32-bit floating point data containing the normalized speckle intensities (ranging from 0 to 1), the intensity values being normalized by the maximal and minimal speckle intensity values in the entire array (or entire recording or part of the recording). 3D speckle images were resampled (custom-made python routine, using VTK-ProbeFilter resampling filter, VTK Visualization Toolkit, Kitware Inc.) from a pyramid-shaped image format into a Cartesian image format with a regular voxel grid structure with grid sizes in the order of 100 × 100 × 100 voxels. The height of the pyramid corresponds approximately to one of the lengths of the box. 2D speckle image sizes were in the order of 500 × 500 pixels and typically reduced in size to about 300 × 300 pixels.

Motion tracking and strain analysis in ultrasound data. Motion within the volumetric (3D, pig) or cross-sectional (2D, rabbit) ultrasound movie data was tracked using custom-made motion-tracking algorithms^{31–33} based on speckle correlation and block-matching. The same algorithm was used for the 2D and 3D

data. Only voxels and pixels showing the heart muscle were considered during the tracking procedure. The image data were segmented into two regions, one showing tissue and one showing the bath (pixel and voxel values: 0 = no tissue, 1 = tissue, ImageJ/Fiji segmentation editor and MATLAB). Ball- (3D) or disk-shaped (2D) blocks or interrogation windows (with diameters of 8–12 voxels and 16–22 pixels) were used to determine shifts of the tissue (within a radius of $r = 4$ voxels and 10 pixels) in the image data and associate the positions of shifted tissue regions in between two frames. The displacements were determined by minimizing the mean quadratic difference (or absolute difference) of two blocks from two consecutive frames. Sub-pixel displacement accuracy was achieved by computing image moments (weight or centroid in 2D or 3D) in the resulting data containing the spatial distribution of mean quadratic differences for given (whole pixel) shifts. The instantaneous displacements between two subsequent frames were then tracked throughout the sequence of images. Despite the noisy speckle image data, tracking was found to detect a coherent displacement field indicating the shifts of the tissue robustly with few outliers. The resulting time-varying, two- or three-dimensional displacement vector fields containing the shifts of the tissue from one (volume) frame to the next were stored and erroneously tracked displacements were determined by computing the deviations of the vectors in a small (ball- or disk-shaped) neighbourhood of pixels (3–5 voxels or pixels diameter) around each voxel or pixel and outliers were removed based on a threshold value. The missing vectors were substituted by vectors interpolated from the remaining surrounding displacement vectors (within 5–7 voxel or pixel diameter). The resulting time-varying vector field of instantaneous displacements was smoothed in space and over time using spatiotemporal filter kernels with a Gaussian profile (radius $r = 12$ –16 voxels and 60–80 pixels, 3–5 frames linear averaging). Scalar-valued measures of deformation (tensor invariants, that is, the tensor trace, principal stretches or principal eigenvalues, red–blue colour code with red indicating contractile and blue indicating tensile strain rates) were computed from the gradient deformation tensor and the Green–Lagrangian strain tensor, which were in turn computed from the smoothed displacement vector field. The strain invariant (tensor trace) was normalized using sliding-window normalization with a w of at least $1 \times$ the period of the dominant frequency and typically not more than 1.5–2.5 periods of the dominant frequency. The resulting time-varying two- or three-dimensional strain rate map was smoothed in space and over time using spatiotemporal filter kernels with a Gaussian profile ($d = 7$ –11 voxels or pixels, 3–5 frames linear averaging). All computational routines for extraction, processing and visualization of the ultrasound imaging data were custom-made routines written in C/C++, VTK and OpenGL.

Computation of phase singularities. From the amplitude of membrane potential, intracellular calcium, surface and intramural or volume strain, the phase was computed. The phase was computed similarly for both experimental and simulation data using the Hilbert transform. Phase singularities were obtained by computing the line integral^{12,13}

$$\oint \nabla \varphi(\mathbf{r}) d\mathbf{s} = 2\pi(p - n) \quad (1)$$

where $\varphi(\mathbf{r})$ is the local phase at a location vector \mathbf{r} and the integral is taken over the closed path s . p and n are the numbers of positively and negatively charged phase singularities enclosed within the path s . equation (1) was used previously for the characterization of vortex dynamics during fibrillation^{6,12,13,34}. The statistics of electrical and mechanical phase singularities were obtained from $n = 3$ pig and $n = 3$ rabbit hearts. Tracking of mechanical phase singularities enabled the reconstruction of the motion and evolution of filaments or phase singular points across space and time. Locations of phase singular points or lines within the moving and deforming medium were stored in the Lagrangian laboratory-coordinate frame.

Computer simulations. Computer simulations were performed using realistic heart geometries as well as bulk-shaped geometries. Actively contracting cardiac tissue supporting nonlinear waves of electrical excitation was simulated using electromechanically coupled reaction–diffusion mechanics numerical models^{35,36}. The three-variable Fenton–Karma model⁵ was used for simulations in realistic, heart-shaped geometries. The two-variable Aliev–Panfilov model³⁷ was used for simulations in cubic-shaped domains. Parameters for the Fenton–Karma model were set to $g_{\text{fi}} = 5.6093115$, $\tau_{\text{r}} = 62.41$, $\tau_{\text{si}} = 63.4124$, $\tau_0 = 8.36$, $\tau_{\text{v}+} = 10.8037$, $\tau_{\text{v}1-} = 40.4284$, $\tau_{\text{v}2-} = 36.5891$, $\tau_{\text{w}+} = 663.715$, $\tau_{\text{w}-} = 132.373$, $u_{\text{c}} = 0.13$, $u_{\text{v}} = 0.05$, $u_{\text{csi}} = 0.45$ with settings for scaling and diffusion properties as previously described⁵. Parameters for the Aliev–Panfilov model were set to $a = 0.05$, $b = 0.5$, $\mu_1 = 0.1$, $\mu_2 = 0.3$, $\varepsilon = 0.01$, as previously described³⁵. Soft-tissue elasticity was simulated using a discrete particle system with controllable elastic anisotropy, as previously described^{32,38}. Bulk-shaped simulation domains consisted of $80 \times 50 \times 80$ hexahedral cells. Realistic, heart-shaped domains were obtained from computerized tomography scans of the rabbit hearts used in the

experiments by segmenting the voxel data and discretizing the segmented heart-shaped domains into tetrahedral volume elements. The simulation grid typically consisted of $120 \times 120 \times 120$ simulation grid cells, which each consisted of up to four tetrahedral cells if containing the segmented elastic medium, less than four cells if containing boundaries of the medium, or no cells if situated outside the medium. Electromechanical coupling was realized by introducing a partial differential equation to model excitation-induced active stress build-up, as previously described^{35,36,39}. The active stress variable was used to modulate the spring constants of the particle system to model active contractions. However, only springs pointing along the local muscle fibre orientation were modulated, introducing an actively contracting preferred orientation in the elasticity, as well as local transverse isotropy of the passive elasticity^{32,38}. The active springs can point into arbitrary directions in the two- or three-dimensional space.

Time-series and signal analysis. Frequency power spectra of electrocardiograms and cumulative frequency power spectra of the 2D and 3D experimental data (voltage, calcium, strain rate) and 2D and 3D simulation data (electrical activity, strain rate) were computed using the fast Fourier transform. First, the fast Fourier transform was applied to the entire time-series with typical durations of 2–10 s, with sampling rates of 85–300 Hz (ultrasound), 250–500 Hz (fluorescence imaging) and 2 kHz (electrocardiogram). Next, its absolute square was computed to yield the frequency power spectrum. For spatially extended data (ultrasound, fluorescence imaging and simulations), the calculation was carried out independently for each pixel or voxel, and the resulting spectra were then averaged across the whole spatial domain to yield the cumulative power spectrum.

Cardiac preparation *ex vivo*. Göttingen minipigs (female, 21–26 months old, 45 ± 3.5 kg) and New Zealand white rabbits (female, 6–12 months old, 2.5–3.5 kg) were anaesthetized (pig: pentobarbital; rabbit: trapanal) and euthanized via an intravenous injection of potassium chloride (KCl). Immediately after euthanasia, cardioplegic solution was injected intravenously for temporary cessation of cardiac activity. Hearts were excised rapidly and inserted into cardioplegic solution for temporary cessation of cardiac activity. Isolated hearts were transferred quickly into a bath filled with warm, oxygenated (95% O₂ and 5% CO₂) Tyrode (15 l distilled water, 130 mM sodium chloride (NaCl), 4 mM KCl, 0.6 mM magnesium chloride hexahydrate (MgCl₂(6H₂O)), 2.2 mM calcium chloride (CaCl₂), 1.2 mM sodium dihydrogen-phosphate NaH₂PO₄(H₂O), 24.2 mM sodium bicarbonate (NaHCO₃) and 12 mM glucose, pH 7.36–7.40) and connected to a retrograde Langendorff-perfusion system (Hugo-Sachs Apparatus). The flow rates of the perfusate were 200 ml min^{−1} and 30 ml min^{−1} for pig or rabbit hearts, respectively, at a perfusion pressure of 50 ± 5 mm Hg. Tyrode was kept at a constant temperature of 37 ± 0.5 °C (custom-made temperature control, Max Planck Institute for Dynamics and Self-Organization) and was constantly reperfused. Any mechanical pressure to the hearts was avoided to prevent compression of the coronary arteries.

Arrhythmia induction and termination, data acquisition. Arrhythmias were induced using rapid pacing via bipolar electrodes or defibrillation paddles, and terminated using a single high-energy cardioversion shock via defibrillation paddles inserted into the bath, if they had not self-terminated, as previously described^{14,15}. Electrical power was supplied by a power amplifier (BOB 100-4M, Kepco Power Supply). Electrocardiograms were recorded using three thin electrodes inserted into the bath and acquired using a data-acquisition system (2.0 kHz sampling rate, MP150, Biopac Systems Inc.).

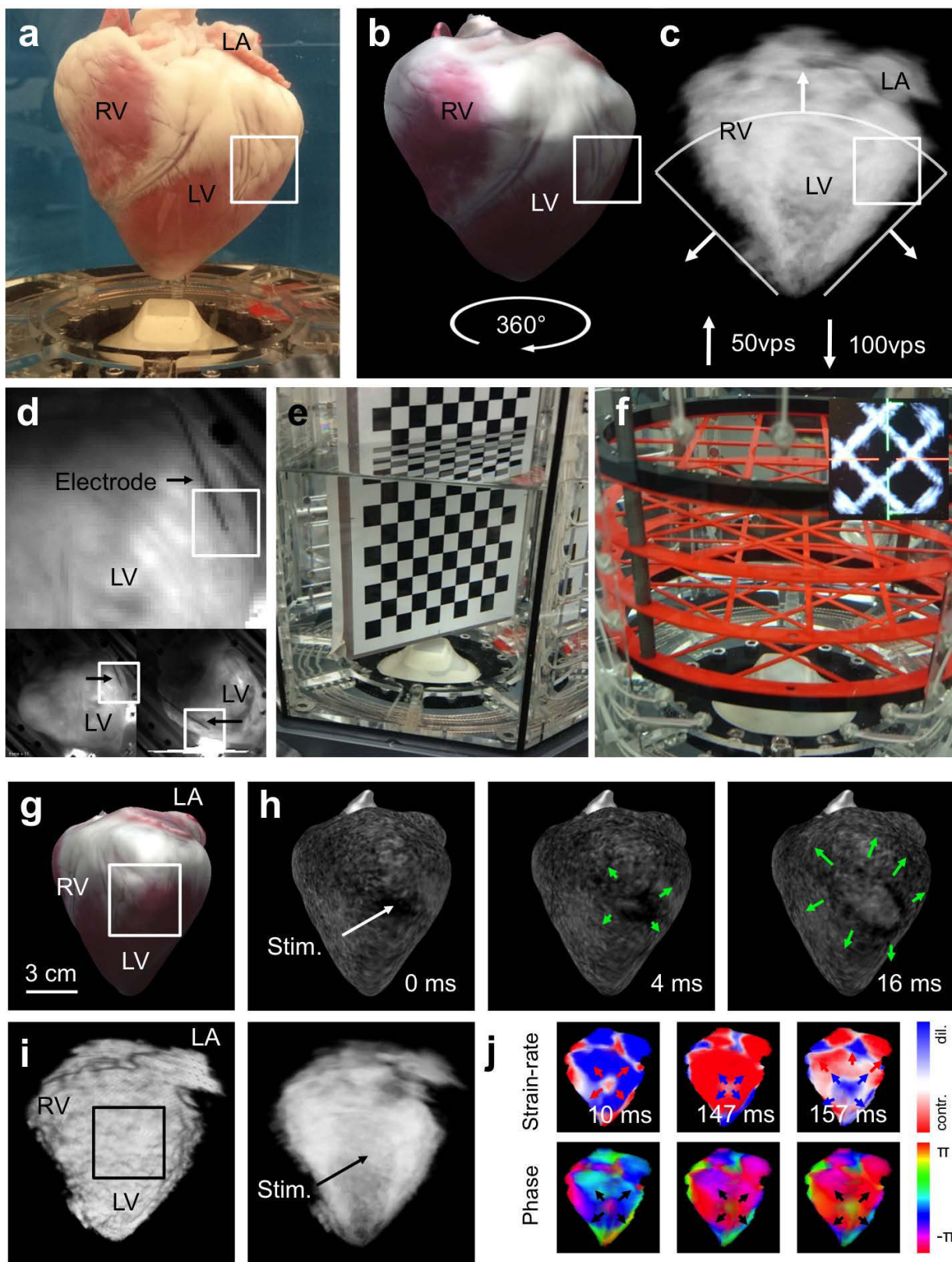
Computerized tomography scans. Micro-computed tomography scans of formalin-fixed rabbit hearts were performed using a commercial computerized tomography scanner (GE CT 120, GE Healthcare) operating at μm spatial resolutions, as previously described¹⁵. Hearts were fixed immediately after the imaging experiments.

Code availability. All custom computer source code used in this study is available from the corresponding author upon reasonable request.

Data availability. The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

- Lee, P. *et al.* Simultaneous measurement and modulation of multiple physiological parameters in the isolated heart using optical techniques. *Pflügers Arch.* **464**, 403–414 (2012).
- Periaswamy, S., *et al.* H. Differential affine motion estimation for medical image registration. In *Proc. SPIE 4119 Wavelet Applications in Signal and Image Processing VIII* <https://doi.org/10.1117/12.408594> (2000).
- Jakubowski, M. & Pastuszak, S. Block-based motion estimation algorithms — a survey. *Opto. Electron. Rev.* **21**, 86–102 (2013).
- Christoph, J. *Intramural Visualization of Scroll Waves in the Heart*. PhD thesis, Univ. Göttingen (2015).
- Huang, Y.-W., Chen, C.-Y., Tsai, C.-H., Shen, C.-F. & Chen, L.-G. Survey on block-matching motion estimation algorithms and architectures with new results. *J. VLSI Signal Process.* **42**, 297–320 (2006).

34. Pathmanathan, P. & Gray, R. A. Filament dynamics during simulated ventricular fibrillation in a high-resolution rabbit heart. *BioMed Res. Int.* **2015**, 720575 (2015).
35. Nash, M. P. & Panfilov, A. V. Electromechanical model of excitable tissue to study reentrant cardiac arrhythmias. *Prog. Biophys. Mol. Biol.* **85**, 501–522 (2004).
36. Panfilov, A. V., Keldermann, R. H. & Nash, M. P. Drift and breakup of spiral waves in reaction–diffusion–mechanics systems. *Proc. Natl Acad. Sci. USA* **104**, 7922–7926 (2007).
37. Aliev, R. R. & Panfilov, A. V. A simple two-variable model of cardiac excitation. *Chaos Solitons Fractals* **7**, 293–301 (1996).
38. Bourguignon, D. & Cani, M.-P. in *Computer Animation and Simulation* (eds Magnenat-Thalmann, N. *et al.*) 113–123 (Springer, 2000).
39. Weise, L. D., Nash, M. P. & Panfilov, A. V. A discrete model to study reaction–diffusion–mechanics systems. *PLoS ONE* **6**, e21934 (2011).
40. Quintanilla, J. G. *et al.* K_{ATP} channel opening accelerates and stabilizes rotors in a swine heart model of ventricular fibrillation. *Cardiovasc. Res.* **99**, 576–585 (2013).
41. Omichi, C., *et al.* Intracellular Ca dynamics in ventricular fibrillation. *Am. J. Physiol. Heart Circ. Physiol.* **286**, H1836–H1844 (2004).
42. Warren, M., Huizar, J. F., Shvedko, A. G., Zaitsev, A. V. Spatiotemporal relationship between intracellular Ca^{2+} dynamics and wave fragmentation during ventricular fibrillation in isolated blood-perfused pig hearts. *Circ. Res.* **101**, e90–e101 (2007).
43. Weiss, J. N., Garfinkel, A., Karagueuzian, H. S., Chen, P. S. & Qu, Z. Early afterdepolarizations and cardiac arrhythmias. *Heart Rhythm* **7**, 1891–1899 (2010).
44. Wang, L., *et al.* Optical mapping of sarcoplasmic reticulum Ca^{2+} in the intact heart: ryanodine receptor refractoriness during alternans and fibrillation. *Circ. Res.* **122**, 1410–1421 (2014).

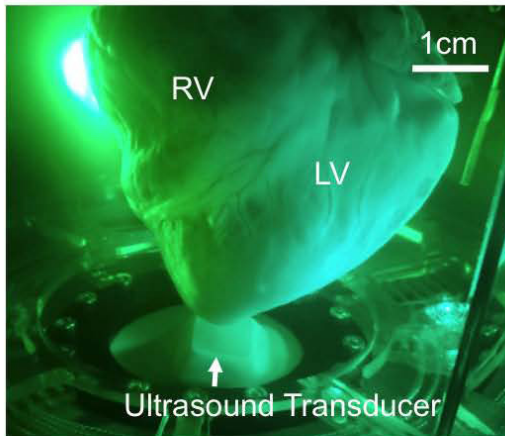
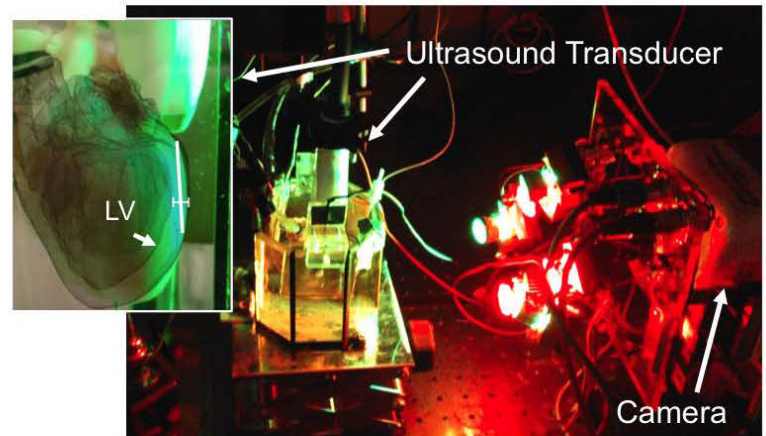
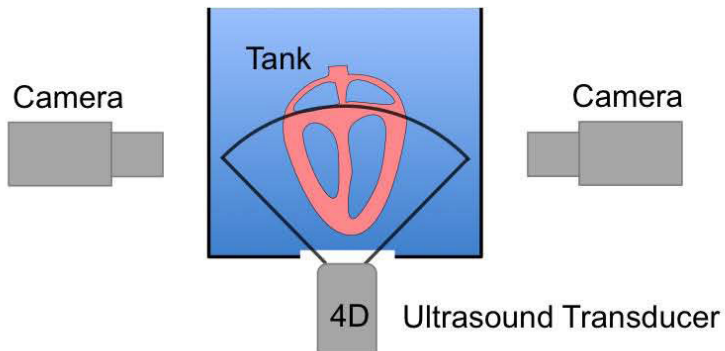
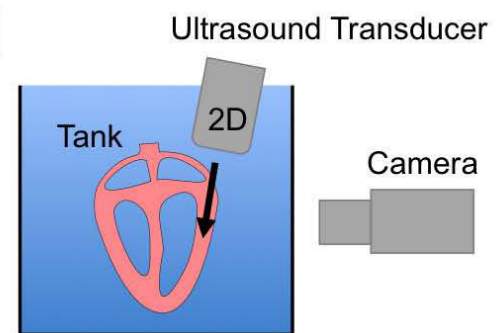
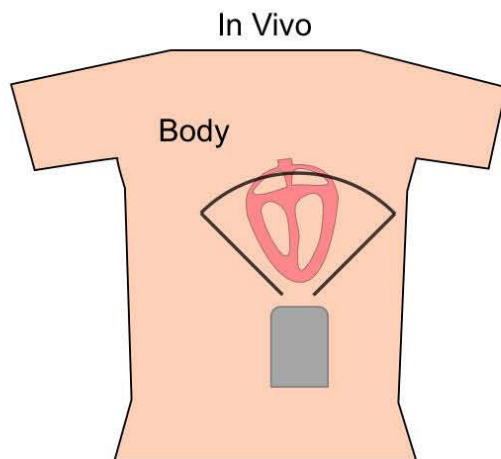
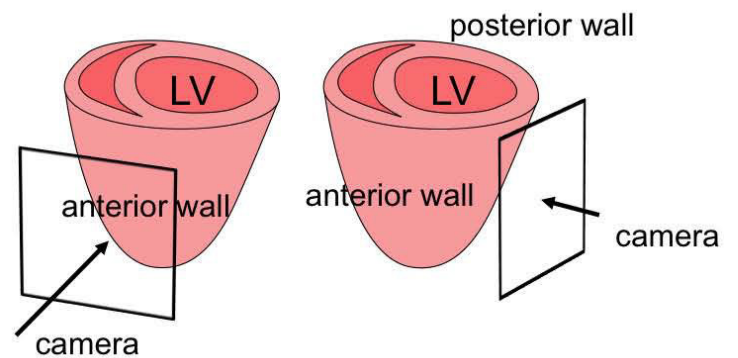


Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | 3D Imaging of electromechanical activity in the heart using simultaneous ultrasound and fluorescence imaging.

a–f, Reconstruction of the heart shape during simultaneous imaging of electromechanical cardiac activity using ultrasound imaging and panoramic optical mapping. **a**, Photograph of heart in the Langendorff setup with ultrasound transducer at the bottom of the bath. LA, left atrium. **b**, 3D optical reconstruction of the surface of the heart obtained by rotating the heart after the experiment. Photorealistic rendering of the surface of the heart on reconstructed 3D surface mesh. **c**, Rendering (transparent) of 3D volumetric ultrasound speckle data showing the contracting heart from a similar perspective. Fan indicates lateral imaging depth of approximately 6–8 cm that was used for high-speed imaging. **d**, Optical mapping video images showing a bipolar electrode touching the epicardium of the left ventricle. The electrode is bend at the top, such that the tip is perpendicular to the electrode (only the tip touches the epicardium). The approximate location of the tip is indicated by the white rectangles in **a–d**, **g**, **i**. The 3D reconstruction allows the determination of the position of the electrode's tip on the heart surface from the two video images. **e**, Optical calibration target used for optical 3D reconstruction of

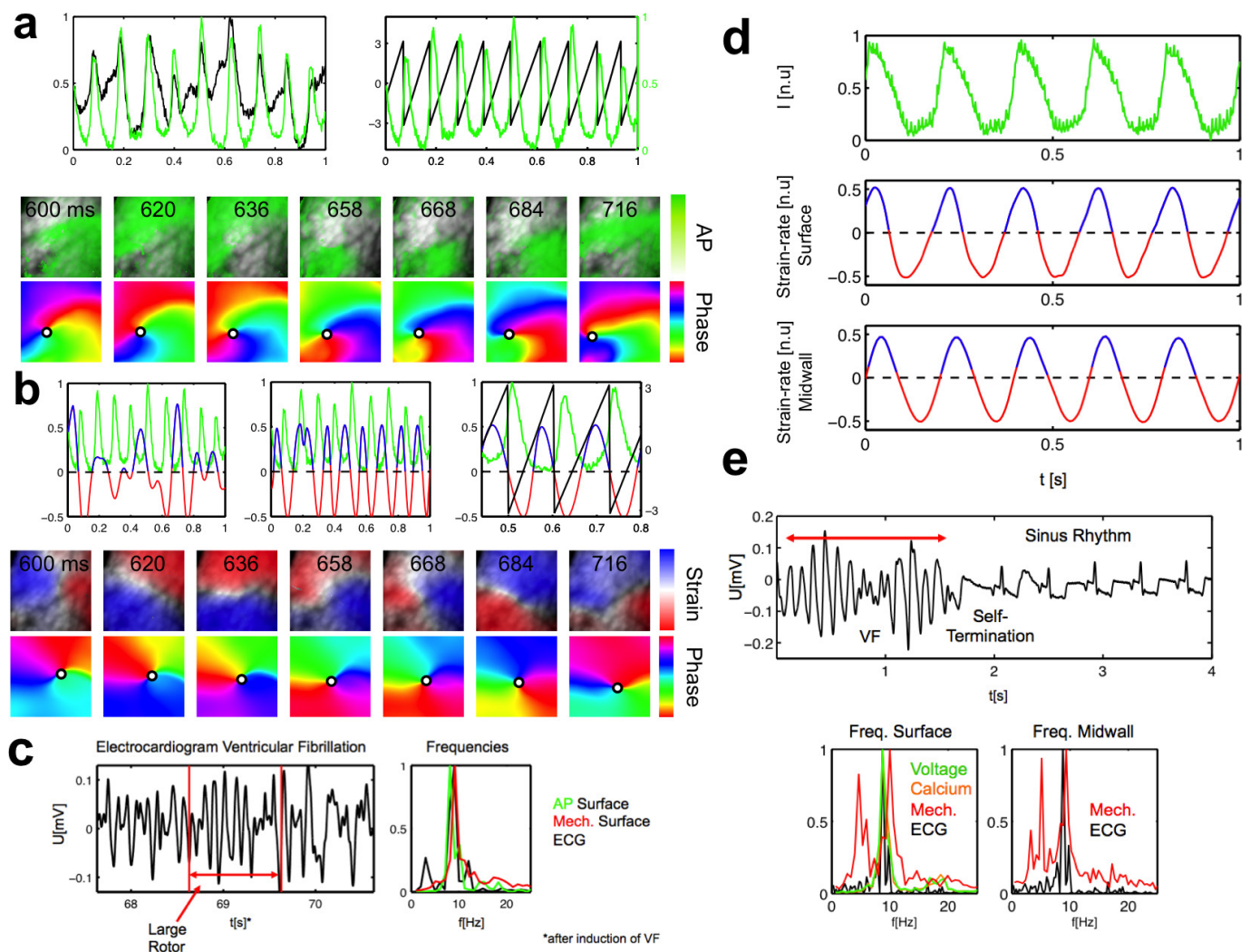
heart surface. **f**, Acoustic calibration target for ultrasound imaging used to determine approximate alignment and position of transducer. The inset shows the corresponding ultrasound speckle image of the calibration target at a particular depth. **g–j**, Focal electromechanical activity in a contracting pig heart mapped with optical mapping and ultrasound during pacing. **g**, Photorealistic rendering of optically reconstructed heart. The heart is rotated towards the left compared to **a**, **b**. **h**, Action potential wave propagating outwards (green arrows) from stimulation (voltage-sensitive mapping using Di-4-ANNEPS, black–white colour code: dark corresponds to depolarizing tissue or the upstroke of the action potential). **i**, 3D ultrasound speckle movie data showing same part of the heart (opaque and transparent rendering). **j**, Elastomechanical activation computed from 3D motion of the tissue acquired with ultrasound. The tissue starts to contract first at the stimulation site. The strain rate and its phase representation pattern exhibit a focal point that coincides with the stimulation site and the focal point that is visible in the optical maps. **a–d**, **g**, **i**, The region in which electrical pacing (stimulation) was applied to the heart is indicated with a rectangle.

a**b****c****d****e****f**

Extended Data Figure 3 | See next page for caption.

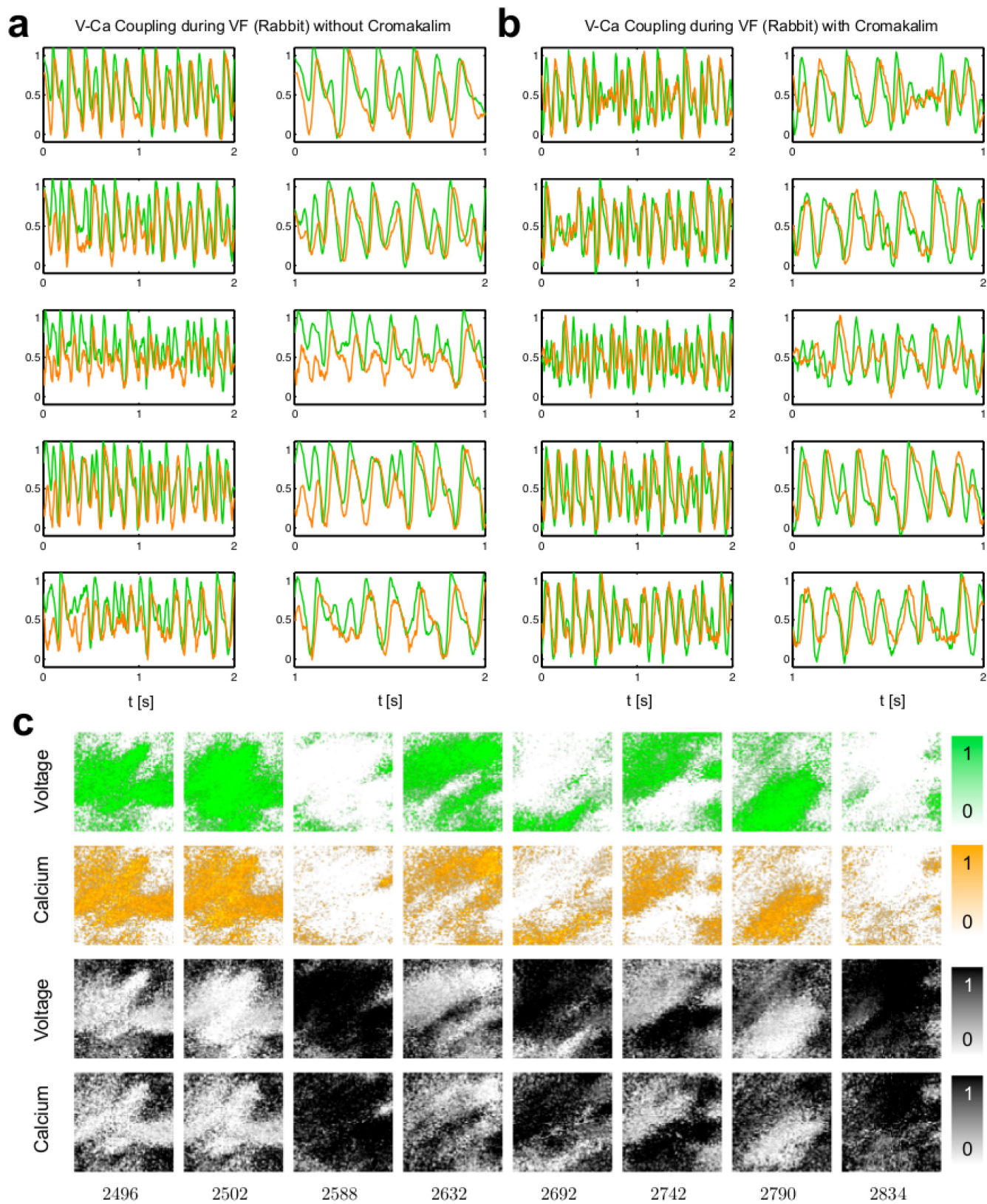
Extended Data Figure 3 | Volumetric and cross-sectional ultrasound imaging and imaging configurations during *ex vivo* and *in vivo* imaging. **a**, Isolated pig heart in imaging setup with 4D volumetric ultrasound transducer scanning from beneath the heart through an acoustic window, together with simultaneous panoramic fluorescence imaging (Fig. 1). 4D ultrasound imaging (Acuson sc2000, Siemens AG) was performed using a two-dimensional matrix phased-array transducer (4Z1c) producing volume frames with a pyramidal shape (51–188 volumes per second, maximum opening angle 90×90 degrees, 6–14 cm lateral depth). The heart is aligned very similarly within the pyramid-shaped field of view as in an examination of a patient *in vivo*. **b**, Isolated rabbit heart in imaging setup with 2D cross-sectional ultrasound transducer scanning from above, together with simultaneous single-camera fluorescence imaging (see also **d** and Supplementary Video 9). **c**, Imaging configuration during 4D ultrasound imaging and panoramic multi-camera optical mapping with four cameras (see also **a**). The optical axis of the cameras and the lateral axis of the ultrasound scanner are approximately perpendicular. **d**, Imaging configuration during 2D ultrasound imaging and single-camera optical mapping (see also **b**). The optical axis of the camera and the lateral axis of the ultrasound scanner are approximately perpendicular. The cross-sectional ultrasound imaging plane is positioned

and aligned tangentially within the ventricular wall underneath and co-planar to the imaged surface. **e**, Imaging configuration during 3D ultrasound imaging *in vivo*. In ultrasound examinations of patients, the human heart is either imaged from outside the body through the chest in a transthoracic echocardiography (TTE) examination, or from within the body in a transoesophageal echocardiography examination using a transoesophageal ultrasound probe. The first measurement configuration provides a clearer picture of the ventricles, whereas the latter measurement provides a clearer picture of the atria. We aimed to mimic the TTE measurement imaging the ventricles using a probe that is routinely used in TTE examinations. In a TTE measurement, the transducer may be positioned on the chest below the heart imaging upwards through two ribs or from underneath the ribs. In one of the various views, the apex of the heart is facing the transducer (apical view), as it is located closest to the transducer array. The atria are located furthest away from the transducer array. The imaging configuration used in our experiment is very similar to this situation (**a**). The size of the human heart is comparable to the size of the pig hearts used in the *ex vivo* experiments. **f**, Frequent imaging configuration during 3D ultrasound imaging with panoramic optical mapping. One camera always imaged the anterior left ventricular wall (similar perspective as in **a**).



Extended Data Figure 4 | Mapping and signal analysis of electromechanical rotor patterns during ventricular fibrillation and tachycardia in isolated Langendorff-perfused hearts. **a**, Mapping of epicardial electrical (voltage) rotor activity during ventricular fibrillation on the contracting surface of a rabbit heart. Left, the time-series (black) obtained from the unstabilized fluorescence maps shows substantial motion artefacts, whereas the time-series of the motion-stabilized maps (green) shows a series of action potentials. Right, from the stabilized time-series, it is possible to compute the phase (black), that is, introducing phase jumps at the upstroke of the action potentials. Bottom, the upper image series shows a counter-clockwise-rotating action potential rotor (green) on the contracting heart surface (Supplementary Video 6). The lower image series shows the corresponding phase representation of the rotor and a phase singularity at the tip of the spiral. **b**, Left, time-series of strain rate obtained from the displacement field. Middle, normalized strain rate (see Methods). Right, zero-crossings of strain rate are used to obtain phase representation. Bottom, rotating wave of dilating (blue) and contracting (red) tissue regions rotating around each other and corresponding phase representation. A circle indicates the position of the mechanical phase singularity. Electrical and mechanical phase singularities are co-localized (Supplementary Video 6). **c**, Electrocardiogram during ventricular fibrillation (same heart as shown in **a**, **b**). Frequency spectra of electrocardiogram (black) and electrical

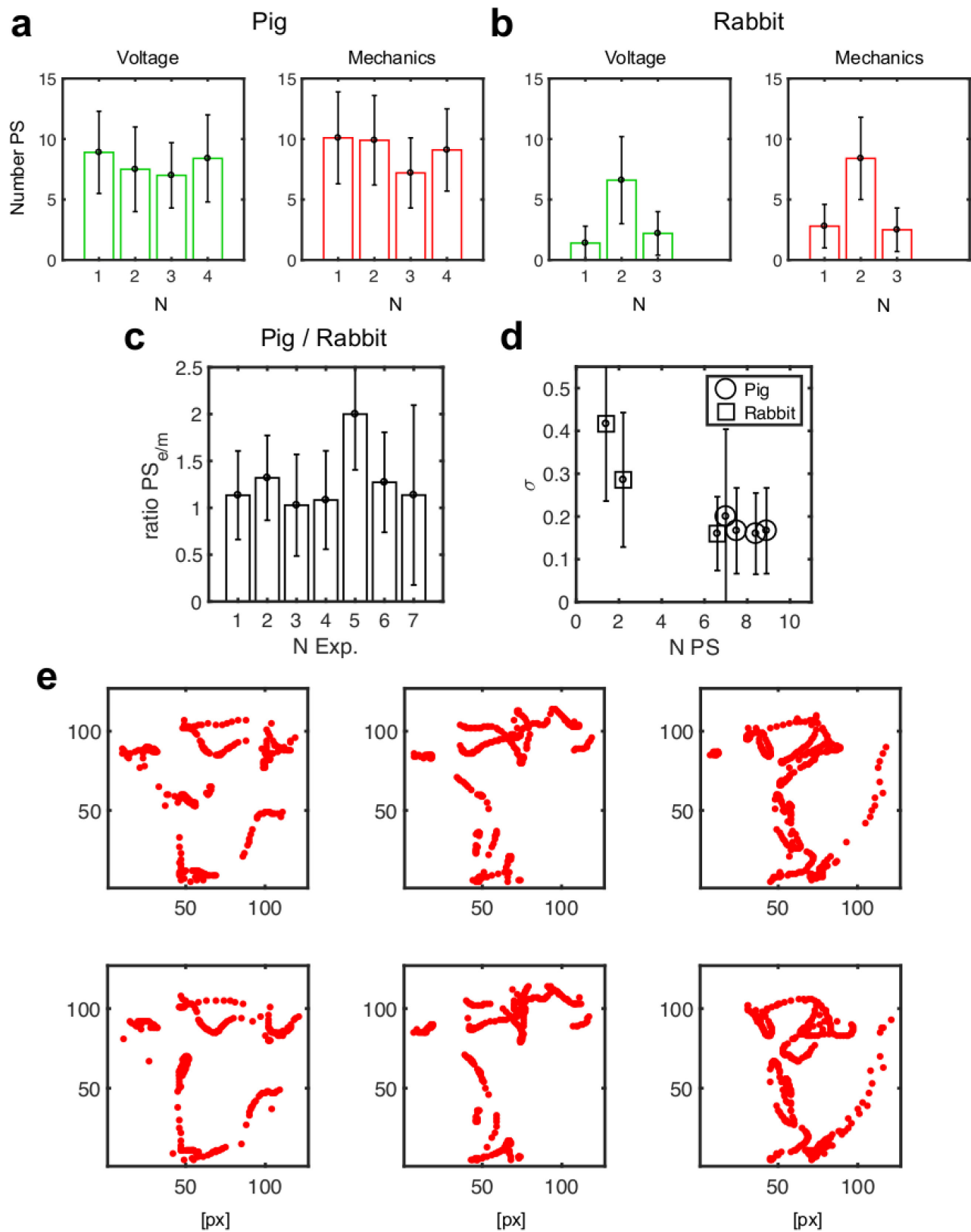
(voltage, green) and elastomechanical (red) patterns measured optically on the epicardial surface of the left ventricular wall (**a**, **b**), showing matching dominant frequencies (9 ± 1 Hz). Frequency spectra of the patterns are cumulative spectra of all measurement sites (Supplementary Video 6). **d**, Electromechanical activity in isolated pig heart during ventricular tachycardia (Fig. 1 and Extended Data Fig. 1b). Top, time-series of action potential activity measured optically using voltage-sensitive optical mapping on anterior left ventricular surface. Middle, time-series of strain rate measured optically in same location as in the top panel on left ventricular surface using fluorescence imaging. Bottom, time-series of strain rate measured within the anterior left ventricular wall using 4D ultrasound imaging. n.u., normalized units. **e**, Electrocardiogram acquired in the isolated rabbit heart during ventricular fibrillation (Fig. 4i). The arrhythmic episode ended by self-termination of arrhythmic activity and transition into sinus rhythm. Frequency spectra of electrocardiogram and electrical (voltage, green) and elastomechanical (red) patterns measured optically on the epicardial surface and elastomechanical (red) patterns measured using ultrasound at midwall inside the left ventricular wall during arrhythmic episode (including data within 2.0 s before self-termination). The dominant frequencies are 9 ± 1 Hz for voltage, calcium and strain. Frequency spectra of the patterns are cumulative spectra of all measurement sites.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Coupled voltage and calcium dynamics mapped during ventricular fibrillation. Voltage (action potential, green, voltage-sensitive staining using Di4-ANBDQPQ) and calcium (intracellular calcium, orange, calcium-sensitive staining using Rhod2-AM) vortex wave activity mapped on ventricular surface of the isolated rabbit heart during ventricular fibrillation using dual-imaging²⁹. Recordings were acquired in non-contracting hearts after administering of blebbistatin (4 ml diluted in 700 ml Tyrode, retrograde perfusion with constant reperfusion). **a, b**, Example time-series ($n = 5$) of voltage–calcium activity chosen from arbitrary sites on the surface. Time-series show closely coupled voltage–calcium wave dynamics with a dominant frequency of 9 ± 1 Hz and a delay of the calcium activity. **a**, Activity on the surface within 2 s long (left) and 1 s long intervals (right) to facilitate viewing of the upstrokes and the delay between voltage and calcium. Recordings acquired before administering (**a**) and after the addition of cromakalim (**b**, 10 μ mol in 1 l Tyrode, retrograde perfusion with constant reperfusion). We did not observe a noticeable difference in the dynamics. **c**, Voltage (action potential, green, voltage-sensitive staining using Di4-ANBDQPQ) and calcium (intracellular calcium, orange, calcium-sensitive

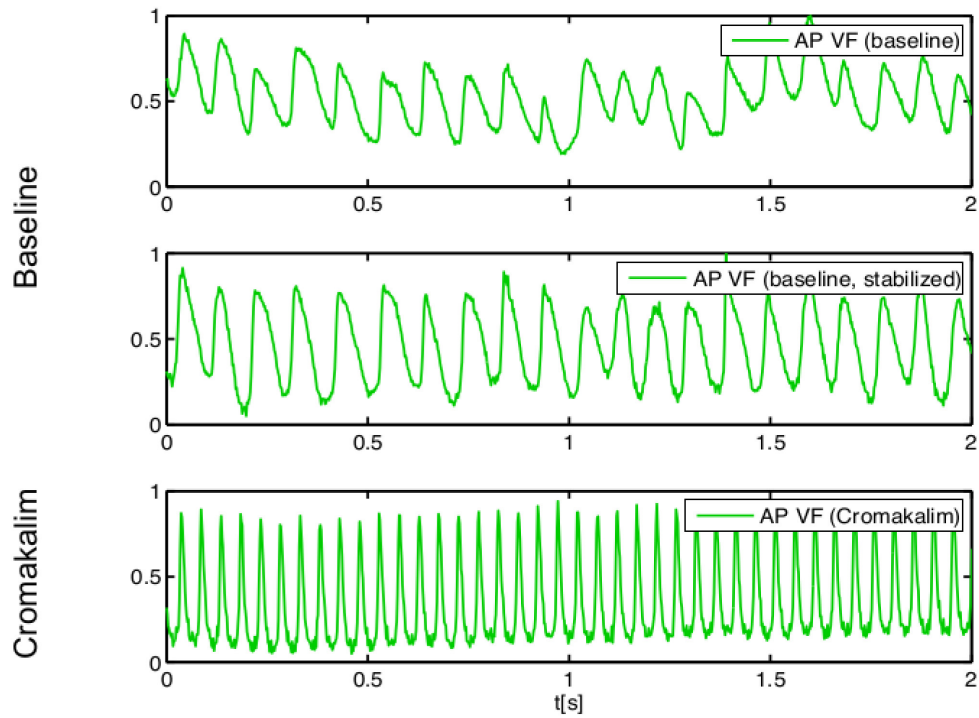
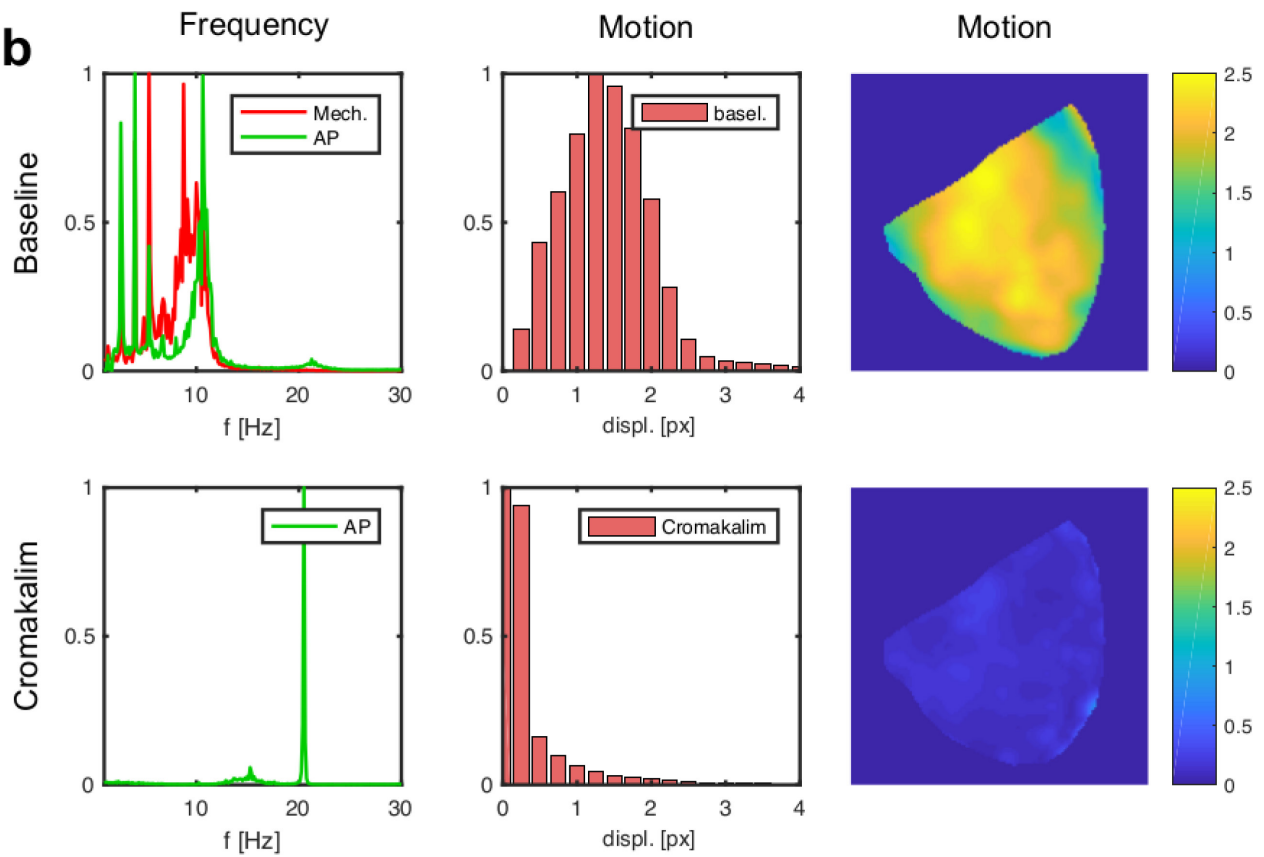
staining using Rhod2-AM) raw movies (black-and-white colour code) mapped on ventricular surface (field of view approximately 1.5×1.5 cm², time in ms) during ventricular fibrillation. The voltage-sensitive data are inverted (dimensionless normalized units (n.u.)), such that bright (or green) areas show depolarized tissue. Bright (or orange) areas in the calcium-sensitive data (dimensionless normalized units (n.u.)), show high intracellular calcium concentrations. The calcium-sensitive images show the same part of the tissue that is shown in the voltage-sensitive images, however, with a delay of 10 ms to compensate for the delay between voltage and calcium and to facilitate analysis of similar wave patterns. The patterns are highly similar and evolve congruently across the surface during ventricular fibrillation. Example image series chosen from a recording with a duration of 20 s. The findings are the same across different recordings. The experiment was repeated twice in $n = 2$ hearts. However, other studies^{41–44} have also shown that in abnormal voltage and calcium regimes the dynamics can have complex interactions and are not always necessarily closely coupled as shown here (see also Supplementary Discussion).



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Characteristics of epicardial electrical and mechanical phase singularities during ventricular fibrillation in pig and rabbit hearts. **a, b**, Mean number of electrical (green) and mechanical (red) phase singularities (PS) during ventricular fibrillation imaged on the epicardial ventricular surface of isolated Langendorff-perfused pig hearts (**a**; four measurements from $n = 3$ hearts) and rabbit hearts (**b**; three measurements from $n = 3$ hearts). The number of phase singularities fluctuates strongly over time (error bars indicate the standard deviation; see also Fig. 4g). Both the electrical and mechanical average numbers of phase singularities similarly reflect different regimes of ventricular fibrillation (**b**). We consistently observed a slightly larger number of mechanical phase singularities (factor 1.2 ± 0.1). The average numbers of phase singularities were computed from 10-s long (500 frames per second) or 20-s long (250 frames per second) recordings with 5,000 video images, including $>10,000$ measurements of phase singularities, the recordings were more than 100 times longer than the average period or lifetime of a rotor. **c**, Ratio of the mean number of electrical and mechanical phase singularities (ratio = number mechanical PS/number electrical PS) during ventricular fibrillation on epicardial ventricular surface of isolated Langendorff-perfused pig hearts (four measurements from $n = 3$ hearts) and rabbit hearts (three measurements from $n = 3$ hearts). The ratio is close to 1 (1.2 ± 0.1) and consistently larger than 1, indicating that more mechanical than electrical phase singularities appear on the surface during ventricular fibrillation. Error bars are large as the number of phase singularities fluctuates strongly (from 0 to approximately 10) over long times (Fig. 4g). Error bars were computed as the standard deviation of the fluctuations of number of phase singularities over time ($>1,000$ samples or time-steps). The ratio (or centre of the plot) was computed from the simple average of the number of phase singularities ($>1,000$ samples or time-steps). **d**, Co-localization factor sigma indicating the precision, with which a mechanical phase singularity describes on average the position of a nearby electrical phase singularity. For large phase singularity numbers ($n_{PS} = 6-10$; right) the precision is about

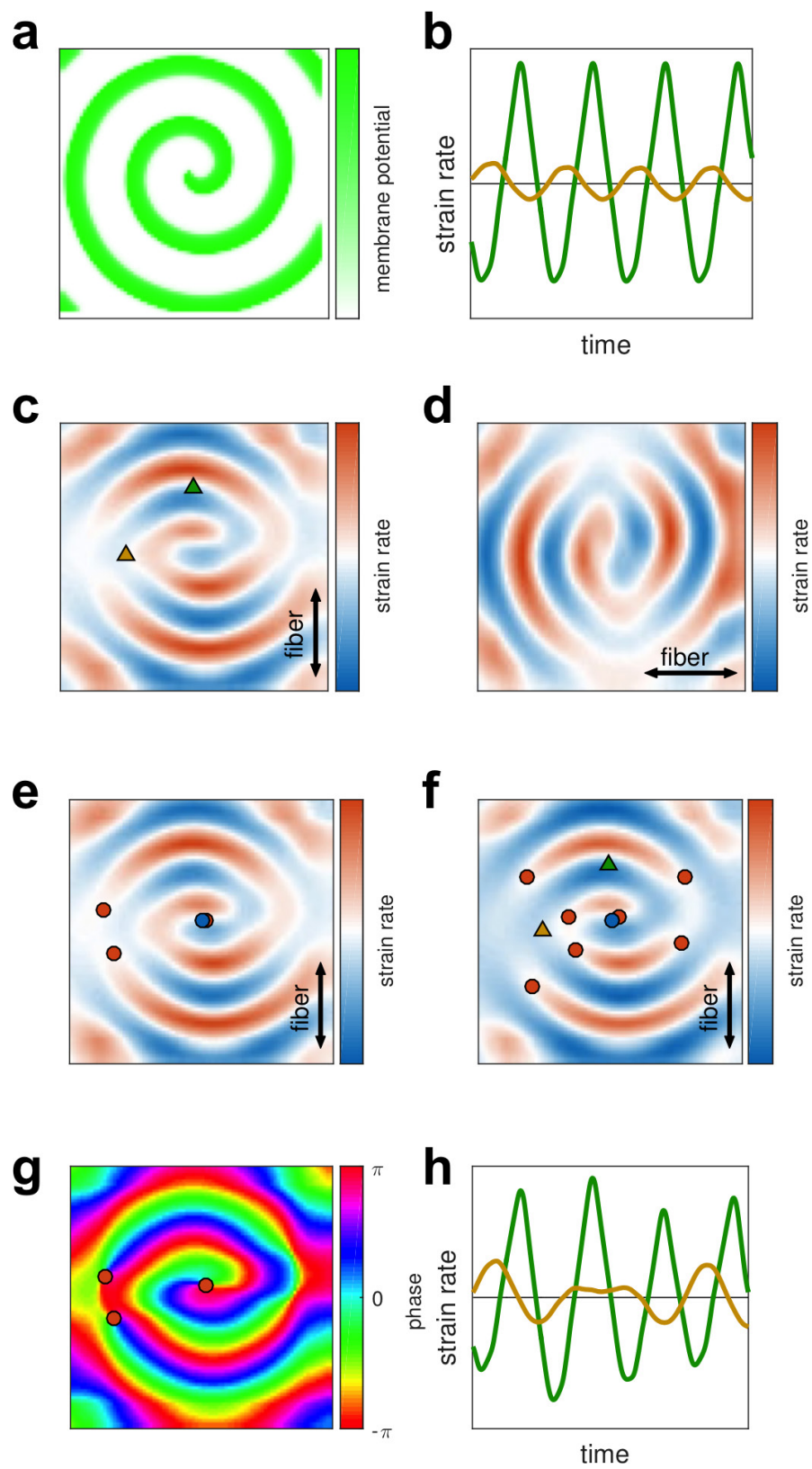
0.2 or 1/5 of the average rotor distance (of 1). In this regime, two phase singularities can clearly be separated from each other (Fig. 4a). For the computation of sigma, the average rotor distances and co-localization distances from the distributions as shown in Fig. 3g and Fig. 4a were used. For smaller phase singularity numbers ($n_{PS} = 1-3$; left) sigma increases, indicating that the precision with which a mechanical phase singularity predicts the location of an electrical phase singularity decreases. The two data points (on the left) describe dynamical regimes with few strongly meandering rotors with linear cores and larger overall deformations of the cardiac muscle (also data point 5 in **c**) during ventricular fibrillation in rabbit hearts (Fig. 3). However, lower sigmas were observed in both pig and rabbit hearts. Lower sigmas or a higher precision is obtained during ventricular fibrillation with a larger number of smaller rotors and a weaker overall deformation of the cardiac muscle. **e**, Trajectories of mechanical phase singularities during ventricular fibrillation on the surface of the rabbit heart. Mechanical phase singularities (red dots) computed individually from voltage-sensitive (top) and calcium-sensitive (bottom) imaging data during multimodal fluorescence imaging (voltage, calcium and strain) with interleaved acquisition of the two channels (250 frames per second per channel = 500 frames per second). The plots show the accumulated mechanical phase singularities during a 400-ms long time interval at 2 s (left), 4 s (centre) and 8 s (right) of a 20-s long recording. The positions and trajectories of the mechanical phase singularities computed from the voltage or the calcium data are identical or at least almost identical throughout time. The corresponding strain rate patterns, from which the phase singularities were computed, are also almost identical or highly similar in each frame over time. The positions and trajectories of the mechanical phase singularities and strain rate patterns are also identical or at least almost identical when only one dye is used and one of the channels does not contain a fluorescent signal. The data demonstrate that the optically derived strain rate patterns are robust, that is, independent of fluorescence-induced image intensity fluctuations.

a**b**

Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Electromechanical dissociation and loss of contractility during very rapid (20 Hz) ventricular fibrillation in pig hearts. a, Example traces of action potential wave activity measured on the left ventricular epicardial surface of a pig heart during ventricular fibrillation using multimodal optical mapping (voltage and contraction). Before the addition of the K_{ATP} channel opener cromakalim (baseline ventricular fibrillation, top), the ventricular muscle exhibits both fibrillatory electrical as well as contractile activity and associated deformations (see also Supplementary Video 12). Owing to the motion of the tissue, the optical traces are superimposed by motion artefacts, visible as modulations of the baseline of the signal. Motion stabilization and motion artefact removal retrieves the action potential wave activity (middle) with substantially reduced motion artefacts. The associated optical maps show spiral vortex wave patterns on the surface of the heart (see also Supplementary Video 12). The dominant frequency of the basal fibrillatory activity is 10.6 ± 1.5 Hz (**b**). After the addition of cromakalim, the dominant frequency of the electrical activity increases markedly (compared to ref. 40) to 20.5 ± 0.5 Hz, see **b**. At the same time, we measured that the heart does not exhibit any visible motion any longer (Supplementary Video 12). Because the contractions of the cardiac muscle vanish during very rapid ventricular fibrillation, optical maps and traces of the action potential wave activity can be retrieved without numerical motion tracking and motion stabilization (bottom). The modulations of the baseline of the signals, which can be seen during baseline ventricular

fibrillation (10 Hz), do not appear at very rapid ventricular fibrillation (20 Hz), indicating that motion is not present (Supplementary Video 12). The data demonstrate a decoupling of the electromechanical wave dynamics during very rapid ventricular fibrillation induced by cromakalim. **b,** Loss of contractile activity during very rapid (20 Hz) ventricular fibrillation. Frequency spectra of electrical action potential wave pattern (green) and mechanical activity (red) measured during baseline ventricular fibrillation (top left, maxima at approximately 10 Hz) and very rapid ventricular fibrillation (bottom left, 20 Hz peak). The dominant frequency of the action potential wave activity approximately doubles from 10 Hz to 20 Hz after the addition of cromakalim, as previously described⁴⁰. At the same time, the contractile motion of the heart vanishes. Correspondingly, a frequency spectrum of the contractile or mechanical activity could not be computed during very rapid ventricular fibrillation. The histograms (middle) show large amplitudes of motion (1–3 pixels displacements of tissue segments within short time spans of 2–3 wave periods) during baseline ventricular fibrillation (top middle) and substantially reduced or vanishing amplitudes of motion (0–0.3 pixels displacements, Supplementary Video 12) during very rapid ventricular fibrillation (bottom middle). The amplitudes of motion were calculated from the displacement data obtained from the motion-tracking procedure. The corresponding maps (top right and bottom right) show the amplitudes of motion measured across the ventricular surface (yellow indicates 2–3 pixels displacement; blue indicates 0 pixels displacement).



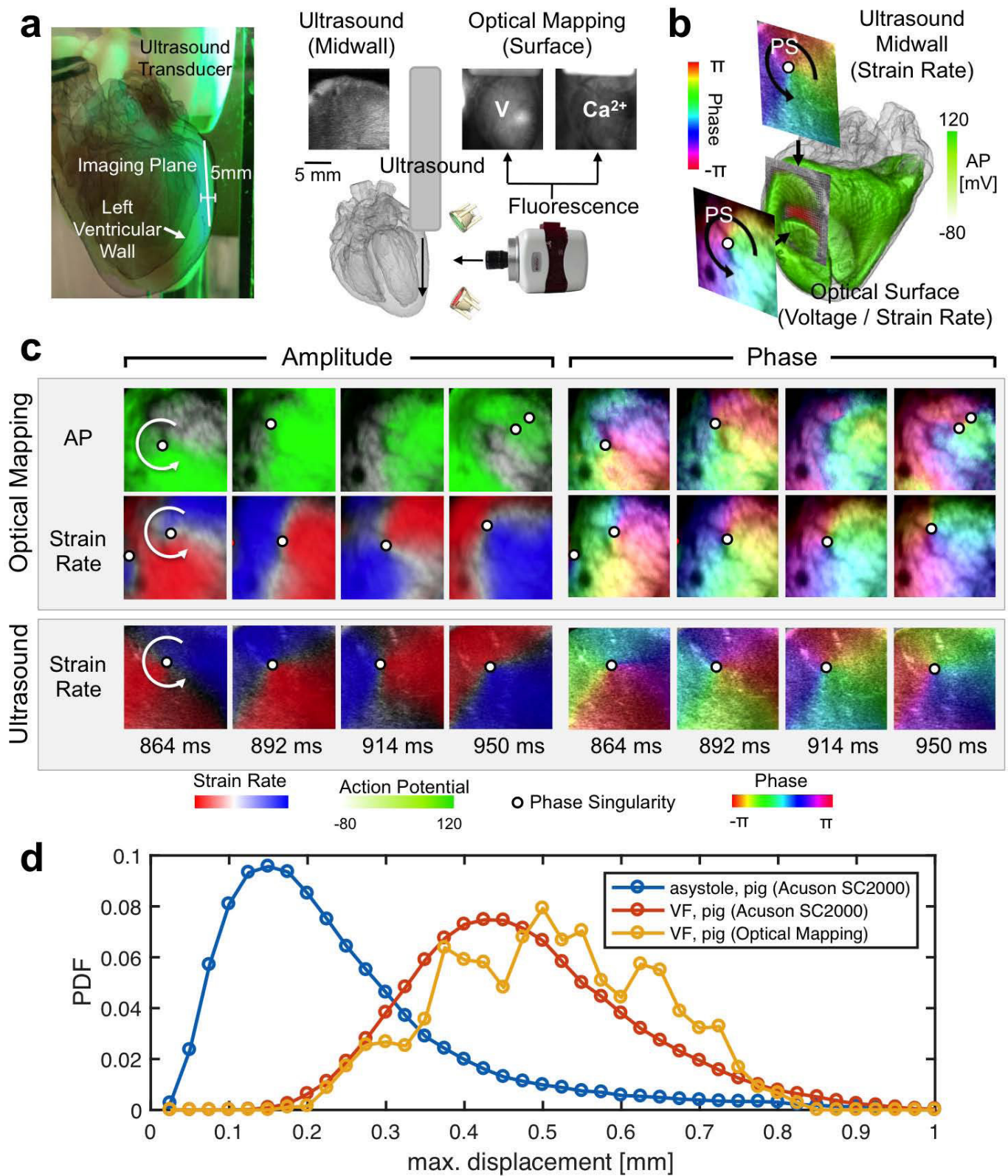
Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Anisotropy of electromechanical rotor patterns in elastic excitable media (computer simulation) and effect of mechanical inhomogeneity onto mechanical phase singularities.

a, b, d. Dependence of local strain rate amplitude and strain rate morphology on the direction of wave propagation relative to muscle fibres. The electrical spiral wave pattern (**a**) and corresponding elastomechanical strain rate patterns (**b, d**) form in two identically prepared simulation domains with differing underlying vertical (**b**) and horizontal (**d**) muscle fibre anisotropy (uniform transverse linear fibre orientation). **c.** The contractions along the fibre orientation produce stronger strain rate amplitudes when the wave propagates along the fibre orientation (green time-series in **b**, sampled from the green triangle) and weaker strain rate amplitudes when the wave propagates perpendicular to the fibre orientation (orange time-series in **b**, sampled from the orange triangle). Overall, the morphology of the strain rate pattern aligns with the fibre orientation and exhibits stronger gradients between dilating and contracting rates of deformation along the fibre orientation than perpendicular to it (polarization). The deformation data were obtained in quasi-2D electromechanical computer simulations with homogeneous active tension development and immediate electromechanical coupling and simulation domains with finite thickness (see Methods).

e–h. Sensitivity of phase singularity detection to external perturbations depending on the direction of wave propagation relative to fibres.

e. Strain rate pattern superimposed with electrical (blue circle) and elastomechanical phase singularities (red circles) in a domain with uniform vertical fibre anisotropy. Perturbations caused by boundary conditions can distort the phase pattern and lead to the detection of spurious mechanical phase singularities, preferentially along the direction that is perpendicular to the fibre orientation. The spurious mechanical phase singularities propagate outwards away from the spiral wave core. **f.** Addition of a weak global strain perturbation exacerbates this effect in the direction that is perpendicular to the fibre orientation. **g.** Phase of strain rate shown in **e** used for the computation of the phase singularities also reveals anisotropy and polarization. **h.** The strong strain rate signal parallel to the fibre orientation (green time-series) is largely unaffected by a small perturbation, whereas, for the same perturbation amplitude, the strain rate signal in the low-amplitude region perpendicular to the fibre orientation is distorted more strongly (orange time-series). The deformation was obtained in simulations identical to those shown in **a–d**. Phase singularities from both the electrical signal and the strain rate were calculated using Hilbert transforms.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Measurement of intramural phase singularity dynamics using 2D ultrasound imaging and tracking of a non-moving heart versus a fibrillating contracting heart in 3D ultrasound data and optical mapping.

a, Schematic of the experimental setup for simultaneous fluorescence and 2D ultrasound imaging in intact, Langendorff-perfused rabbit hearts. Left, ultrasound imaging (Vevo 2100, VisualSonics Inc.) with the cross-sectional echocardiographic imaging plane aligned tangentially inside the left ventricular wall (white line). Imaging plane facing fluorescence imaging camera (Extended Data Fig. 3d). Right, action potential, calcium and strain imaging setup for fluorescence imaging. **b**, Schematic of imaging configuration with action potential wave (phase) imaged on the surface of the heart and rotating mechanical pattern (phase) imaged inside the ultrasound cross-section inside the ventricular wall. The rendering shows a scroll wave (green, computer simulation). **c**, Optical mapping shows a counter-clockwise rotating action potential wave (AP, green) on the surface of a rabbit heart during ventricular fibrillation (Supplementary Video 9). The electrical vortex is associated with a counter-clockwise rotating pattern of contractile (red) and tensile (blue) rates of strain observed optically with fluorescence imaging on the surface. The mechanical deformation observed in B-mode ultrasound imaging in an imaging plane located beneath the imaged surface and aligned approximately parallel or co-planar to the epicardium shows a phase singularity, corresponding to a counter-clockwise rotating wave in mechanical deformation (Supplementary Video 9). The analysis of the electrical and mechanical vortices reveals co-existing phase singularities (indicated by white circles) on the surface and inside the ventricular wall. **d**, Tracking of the non-moving heart compared to a fibrillating contracting heart in 3D ultrasound and fluorescence imaging data. Distributions showing magnitudes of tracked displacement vectors resulting from motion tracking of the 3D ultrasound data (blue and red curves) and motion tracking during optical imaging (yellow curve). Influences by speckle noise or possible residual motion onto the motion tracking and motion analysis during ultrasound imaging are minimal in the Langendorff setup. Motion is not visible (Supplementary Video 13, left) when the heart does not contract. In comparison, the small contractions and deformations during ventricular fibrillation can clearly be observed (Supplementary Video 13, right) and can also be detected using motion-tracking algorithms (red curve, 3D ultrasound; yellow curve, optical mapping). The maximum of tracked displacements for the asystolic heart is approximately 0.15 mm. By contrast, during fibrillation, the maximum

displacements are clearly shifted towards larger values with a maximum at 0.45 mm. The magnitudes of the tracked displacements during ventricular fibrillation (red curve) are confirmed by the optical measurement performed at the same time (yellow curve). Statistical analysis (two-sample Kolmogorov–Smirnov test) rejects the null hypothesis that the ultrasound-based measurements of tissue displacement during asystole and ventricular fibrillation are from the same continuous distribution at a 1% significance level. For the ultrasound data, the distributions were obtained from 274,444 voxels (asystole) and 165,558 voxels (fibrillation), for which the displacements were tracked in between consecutive volume frames (with $\Delta t = 1$ or $\Delta t = 2$ or $\Delta t = 3$) for the entire video sequence. Then only the maximal displacement in each voxel was stored and considered for the distributions to emphasize the influence of noise. Note in this context that in the asystolic case, tracking yields small displacement magnitudes due to measurement noise resulting presumably from both the algorithm and speckle noise. For a better comparison, the Δt or the temporal distance between volume frames between which motion was detected was adjusted. The asystolic data were imaged at volume rates of 62 volumes per second, whereas the fibrillation data was imaged at volume rates of 91 volumes per second. For the asystolic data (left), displacements were computed in between one and the second next volume frame ($\Delta t = 2$), resulting in an effective volume rate of approximately 30 volumes per second. However, the distribution (left) remained unchanged with $\Delta t = 1$, $\Delta t = 2$ or $\Delta t = 3$ frames, indicating that the tissue is static and does not exhibit motion. For the fibrillation data (right), displacements were computed in between one and the third next volume frame ($\Delta t = 3$), resulting also in an effective volume rate of approximately 30 volumes per second. With $\Delta t = 1$ or $\Delta t = 2$ frames, the distribution also remains clearly distinguishable from the non-moving distribution and retains displacement magnitudes well over 0.5 voxels. Note that in this particular fibrillation dataset, the overall strength of motion and deformation is relatively small compared to other datasets (for instance the one shown in Fig. 1). In the optical data, displacements were computed from a frame at time t to a frame at time $t + 32$ ms ($\Delta t = 16$ frames) to achieve a corresponding measurement of tissue displacements at imaging speeds of approximately 30 frames per second and to ensure comparability of the data. The data and the clearly distinguishable visual appearance of the non-moving and fibrillating heart (Supplementary Video 13) demonstrate that the spatiotemporal resolution of the 3D ultrasound imaging is sufficient to resolve elastomechanical deformation patterns during tachyarrhythmias.

Extended Data Table 1 | Statistics of electrical and mechanical phase singularity dynamics on the surface of fibrillating, contracting hearts ($n=6$)

Exp.	# PS V	# PS M	Ratio M/V	Distance V-M [mm]	Dist. PS V [mm]	Dist. PS M [mm]	Ratio Distance V-M
Pig							
1	8.9 ± 3.4	10.1 ± 3.8	1.1	3.4 ± 1.7	20.6 ± 6.9	20.6 ± 6.9	0.17
2	7.5 ± 3.5	9.9 ± 3.7	1.3	3.4 ± 1.7	20.6 ± 6.9	20.6 ± 6.9	0.17
3	7.0 ± 2.7	7.2 ± 2.9	1.0	4.3 ± 4.3	21.5 ± 4.3	21.5 ± 4.3	0.20
4	8.4 ± 3.6	9.1 ± 3.4	1.1	3.4 ± 1.7	20.6 ± 6.9	20.6 ± 6.9	0.17
Rabbit							
5	1.4 ± 1.4	2.8 ± 1.8	2.0	2.9 ± 1.2	5.9 ± 1.2	7.0 ± 1.2	0.50
6	6.6 ± 3.6	8.4 ± 3.4	1.3	0.9 ± 0.5	5.3 ± 1.2	5.9 ± 1.2	0.17
7	2.2 ± 1.8	2.5 ± 1.8	1.1	2.3 ± 1.2	7.6 ± 1.9	8.2 ± 1.9	0.30

Measurements obtained by fluorescence imaging of the left ventricular surface of fibrillating, contracting pig ($n=3$) and rabbit ($n=3$) hearts. The fourth dataset (in the pig) was obtained in the same heart as the third heart, but in a different location (posterior instead of anterior left ventricular wall). Each data point corresponds to a recording, in which many hundreds of measurements were performed throughout the sequence of video images. The table shows similar numbers of electrical (V) and mechanical (M) phase singularities across hearts in the same species and similar ratios (M/V) across hearts and species, respectively. The measurements were obtained from several thousand video frames. In addition, the distances between electrical and mechanical (V-M) phase singularities are smaller than between equal (electrical–electrical or mechanical–mechanical) phase singularities across hearts and species, respectively. The data support the hypothesis that, on average, each electrical phase singularity can be associated with a nearby mechanical phase singularity. The average nearest neighbour distance between an electrical phase singularity and its corresponding mechanical phase singularity is small compared to or a fraction of the distance to the next electrical or mechanical phase singularity, respectively (in almost all cases the fraction is about 0.2).

Hepatocyte-secreted DPP4 in obesity promotes adipose inflammation and insulin resistance

Devram S. Ghorpade¹, Lale Ozcan¹, Ze Zheng¹, Sarah M. Nicoloso², Yuefei Shen², Emily Chen^{3,4}, Matthias Blüher⁵, Michael P. Czech² & Ira Tabas^{1,6}

Obesity-induced metabolic disease involves functional integration among several organs via circulating factors, but little is known about crosstalk between liver and visceral adipose tissue (VAT)¹. In obesity, VAT becomes populated with inflammatory adipose tissue macrophages (ATMs)^{2,3}. In obese humans, there is a close correlation between adipose tissue inflammation and insulin resistance^{4,5}, and in obese mice, blocking systemic or ATM inflammation improves insulin sensitivity^{6–8}. However, processes that promote pathological adipose tissue inflammation in obesity are incompletely understood. Here we show that obesity in mice stimulates hepatocytes to synthesize and secrete dipeptidyl peptidase 4 (DPP4), which acts with plasma factor Xa to inflame ATMs. Silencing expression of DPP4 in hepatocytes suppresses inflammation of VAT and insulin resistance; however, a similar effect is not seen with the orally administered DPP4 inhibitor sitagliptin. Inflammation and insulin resistance are also suppressed by silencing expression of caveolin-1 or PAR2 in ATMs; these proteins mediate the actions of DPP4 and factor Xa, respectively. Thus, hepatocyte DPP4 promotes VAT inflammation and insulin resistance in obesity, and targeting this pathway may have metabolic benefits that are distinct from those observed with oral DPP4 inhibitors.

In obese mice and humans, a pathway in hepatocytes involving Ca^{2+} -calmodulin-dependent protein kinase II (CaMKII) contributes to both excessive hepatic glucose production and impaired hepatic insulin signalling^{9–11}. Inhibition of this pathway by hepatocyte-specific deletion of CaMKII γ in diet-induced obese (DIO) mice (*Camk2g^{fl/fl}* mice infected with AAV8-TBG-*cre*, hereafter referred to as H-CaMKII(KO) mice, improves glucose and insulin tolerance^{9,10}. Notably, the VAT of these mice had fewer ATMs in crown-like structures (CLS) and decreased *Adgre1* mRNA (which encodes the F4/80 glycoprotein) in comparison to VAT of control *Camk2g^{fl/fl}* DIO mice (Extended Data Fig. 1a). There was also decreased infiltration of inflammatory Ly6C^{hi} monocytes and lower inflammatory cytokine expression in the VAT of H-CaMKII(KO) mice, with no significant change in blood monocyte number, plasma IL6 or TNF α , or liver inflammation (Extended Data Fig. 1b–e). ATF4 expression is regulated downstream of CaMKII signalling and is decreased in the hepatocytes of H-CaMKII(KO) DIO mice¹⁰. Restoring ATF4 expression in the livers of mice deficient in hepatocyte CaMKII using an adenovirus vector restored VAT inflammation without affecting liver inflammation (Extended Data Fig. 1f, g). Conversely, hepatocyte-specific deletion of ATF4 in DIO mice lowered VAT inflammation (Extended Data Fig. 1h). Therefore, we hypothesized that, in obesity, activation of CaMKII and ATF4 in hepatocytes induces the secretion of a circulatory factor (a ‘hepatokine’) that promotes VAT inflammation.

To test this hypothesis, we developed an *ex vivo* assay in which cells from the stromal vascular fraction (SVF) of VAT are incubated with plasma from the above mouse models, and *Mcp1* (also known as *Ccl2*)

and *Il6* mRNA are quantified as markers of inflammation. Because VAT macrophages are a likely target of the putative hepatokine, we predicted that cells from the SVF of VAT from obese mice, which harbour more of these macrophages^{4,5} (Extended Data Fig. 2a), would show an increased response in this assay in comparison to SVF cells from lean mice. Plasma from obese mice would also be expected to evoke a more potent response in the assay than plasma from lean mice. Consistent with this hypothesis, plasma from DIO mice induced higher expression of both *Mcp1* and *Il6* mRNA than plasma from lean mice in SVF cells from DIO mice, but did not have this effect on SVF cells from lean mice. Furthermore, the inflammatory activity of plasma from DIO mice was due to the macrophage component of SVF cells (Fig. 1a and Extended Data Fig. 2b). DIO mouse plasma was also able to induce *Mcp1* expression in peritoneal and bone marrow-derived macrophages (BMDMs; Extended Data Fig. 2c). Most importantly, plasma from H-CaMKII(KO) DIO mice was less able to induce *Mcp1* expression in SVF than plasma from either wild-type DIO mice or H-CaMKII(KO) mice in which hepatic ATF4 expression was restored (Fig. 1b). These data suggest that, in obesity, activation of hepatocyte CaMKII and ATF4 induce one or more secretory factors that promote VAT-macrophage inflammation.

The *Mcp1*-inducing factor in DIO mouse plasma was heat labile (Extended Data Fig. 2d), suggesting that it may be a protein. We performed size-exclusion gel-filtration fast protein liquid chromatography (FPLC) with DIO mouse plasma (Extended Data Fig. 2e) and tested eluted fractions for their ability to induce *Mcp1* expression in SVF cells. There was a peak of activity in three fractions in the 125–200 kDa range (Extended Data Fig. 2f). The active fraction F44 and inactive fractions F42 and F46 were analysed by liquid chromatography–tandem mass spectrometry (LC–MS/MS) to identify peptides that were more abundant in F44 than in F42 and F46 (Supplementary Table 2a). Peptides corresponding to DPP4 (CD26) met these criteria (Extended Data Fig. 2g). DPP4 is a dipeptidyl protease that can exist as either a cell membrane protein or a soluble plasma protein¹². DPP4 concentration correlates with body mass index and insulin resistance in humans^{13,14}, and we found that DPP4 activity was higher in the plasma of DIO mice than in that of lean mice (Extended Data Fig. 2h).

Consistent with a role for DPP4, the *Mcp1*- and *Il6*-inducing activity of DIO mouse plasma could be suppressed by an inhibitor of DPP4 (Fig. 1c and Extended Data Fig. 2i). Furthermore, DPP4 activity in plasma and *Dpp4* mRNA expression in liver, but not in VAT, correlated exactly with VAT inflammation *in vivo* and plasma SVF-inflammatory activity *ex vivo* in wild-type and H-CaMKII(KO) mice as well as in H-CaMKII(KO) mice in which hepatic ATF4 expression was restored (Fig. 1d and Extended Data Fig. 2j). Moreover, restoration of hepatic DPP4 expression in H-CaMKII(KO) mice restored DPP4 activity in plasma and VAT inflammation without changing body weight (Extended Data Fig. 3a, b). Consistent with the role of ATF4, we

¹Department of Medicine, Columbia University Medical Center, New York, New York 10032, USA. ²Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ³Proteomics Shared Resource in the Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, New York 10032, USA. ⁴Herbert Irving Comprehensive Cancer Center Proteomics Shared Resource, Columbia University Medical Center, New York, New York 10032, USA. ⁵Department of Medicine, University of Leipzig, Leipzig 04103, Germany. ⁶Department of Pathology & Cell Biology and Department of Physiology, Columbia University Medical Center, New York, New York 10032, USA.

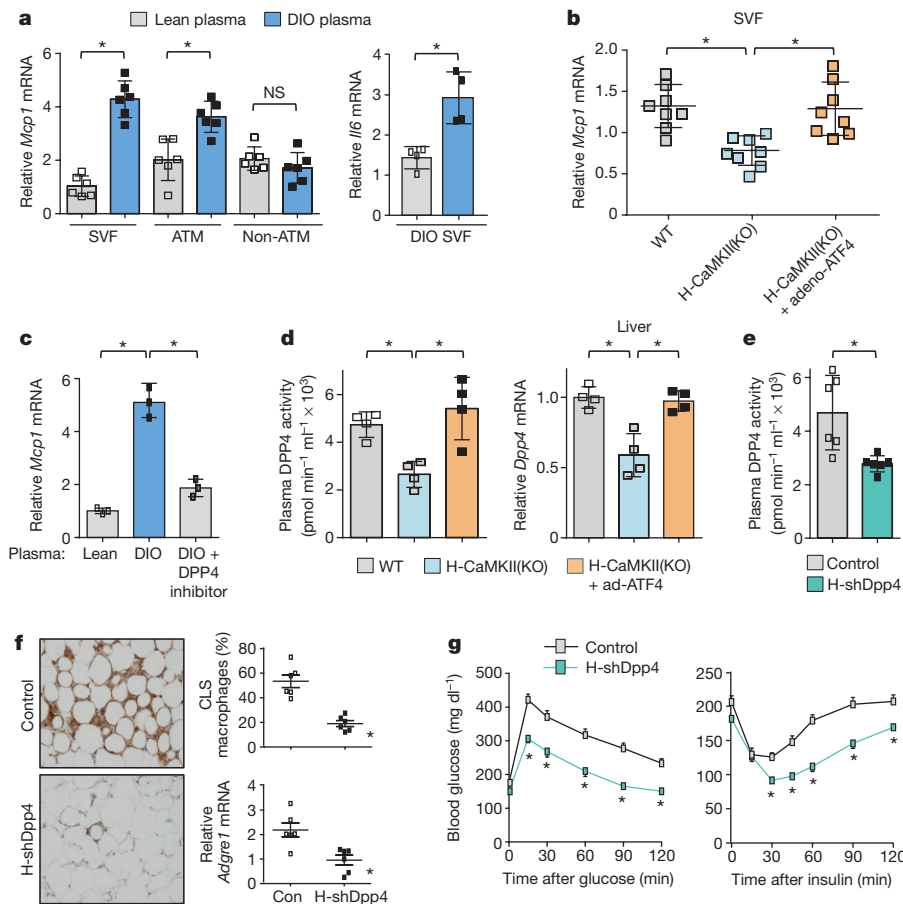


Figure 1 | Hepatocyte-specific DPP4 silencing suppresses VAT inflammation and improves insulin sensitivity in DIO mice. **a**, VAT SVF cells from DIO mice, and ATMs and non-ATMs isolated from SVF, were incubated for 4 h with medium containing 10% (vol/vol) plasma from lean or DIO mice and then assayed for *Mcp1* or *Il6* mRNA ($n = 4-6$ mice per group; $*P < 0.05$ by two-tailed Student's *t*-test). **b**, VAT SVF cells from DIO mice were incubated with plasma from the three DIO models described in Extended Data Fig. 1f, g and then assayed for *Mcp1* mRNA. WT, wild type. $n = 8$ mice per group; mean \pm s.e.m.; $*P < 0.05$ by one-way analysis of variance (ANOVA). **c**, SVF cells from DIO mice were incubated with lean or DIO mouse plasma that was pre-treated for 1 h with or without $10 \mu\text{M}$ DPP4 inhibitor KR62436 and then assayed for *Mcp1* mRNA ($n = 3$

technical replicates; mean \pm s.e.m.; $*P < 0.05$ by one-way ANOVA). **d**, The three groups of mice from Extended Data Fig. 1f, g were assayed for plasma DPP4 activity and hepatic *Dpp4* mRNA ($n = 4$ mice per group; mean \pm s.e.m.; $*P < 0.05$ by one-way ANOVA). **e-g**, Sixteen-week-old mice previously fed the DIO diet for 10 weeks were injected intravenously with AAV8-H1-shDpp4 (H-shDpp4) or control AAV8-H1 vector (control; or con, in **f**). The mice were analysed after four weeks as follows. **e**, Plasma DPP4 activity. **f**, CLS macrophages and *Adgre1* mRNA in VAT, with representative images of F4/80-stained VAT. **g**, Blood glucose after intraperitoneal glucose or insulin ($n = 6$ mice per group; mean \pm s.e.m.; $*P < 0.05$ by two-tailed Student's *t*-test).

found an ATF4-consensus site in exon 1 of the *Dpp4* gene. Chromatin immunoprecipitation (ChIP) analysis revealed lower levels of ATF4 occupancy at this site in the livers of H-CaMKII(KO) DIO mice versus those of wild-type DIO or H-CaMKII(KO) DIO mice in which hepatocyte ATF4 expression was restored (Extended Data Fig. 3c).

We infected DIO mice with an adeno-associated virus 8 (AAV8) that encoded a short hairpin RNA (shRNA) construct targeting mouse *Dpp4* driven by the H1 promoter. AAV8-H1-shDpp4 specifically silences DPP4 in hepatocytes but not in VAT or other tissues (Extended Data Fig. 3d, e); we refer to the vector as H-shDpp4 to reflect its hepatocyte specificity. H-shDpp4-treated DIO mice had significantly decreased plasma DPP4 activity (Fig. 1e), indicating that hepatocytes are a source of circulating DPP4 in obesity. H-shDpp4-treated DIO mice also had decreased numbers of CLS macrophages in the VAT and decreased levels of inflammatory cytokine mRNAs in VAT and VAT ATMs (Fig. 1f and Extended Data Fig. 4a, b). However, these mice did not differ from control vector-treated DIO mice with regard to inflammation in inguinal or brown fat, plasma cytokines, body weight, food intake or liver or adipose tissue weight (Extended Data Fig. 4c-g). H-shDpp4-treated DIO mice also exhibited improved glucose homeostasis and increased insulin-induced AKT phosphorylation (p-AKT) in VAT and

liver (Fig. 1g and Extended Data Fig. 4h, i). Unlike DPP4 inhibition¹⁵, silencing of hepatocyte DPP4 did not increase levels of active plasma incretins (Extended Data Fig. 4j). Similar data were obtained when *ob/ob* mice were treated with the H-shDpp4 (Extended Data Fig. 5a-h). By contrast, treatment of lean mice with the H-shDpp4 had no effect on metabolism (Extended Data Fig. 5i-m), and DPP4 silencing did not improve insulin signalling in palmitate-treated primary hepatocytes (Extended Data Fig. 6a). Moreover, consistent with the idea that adipocyte-derived non-esterified fatty acids (NEFA) may contribute to inflammation-induced insulin resistance¹⁶, H-shDpp4 treatment lowered plasma NEFA in obese but not in lean mice (Extended Data Fig. 6b). These combined data suggest that the improvement in metabolism in DIO mice treated with H-shDpp4 is caused by suppression of VAT inflammation, which improves insulin signalling in hepatocytes and VAT as a secondary effect rather than via a direct effect on insulin signalling in hepatocytes.

Oral DPP4 inhibitors do not lower plasma insulin in insulin-resistant, hyperinsulinaemic subjects¹⁷⁻²². A recent study suggested that in obese mice, oral DPP4 inhibitors function by inhibiting gut endothelial DPP4²³. We tested the hypothesis that oral DPP4 inhibition and hepatocyte DPP4 silencing have fundamentally different effects on adipose

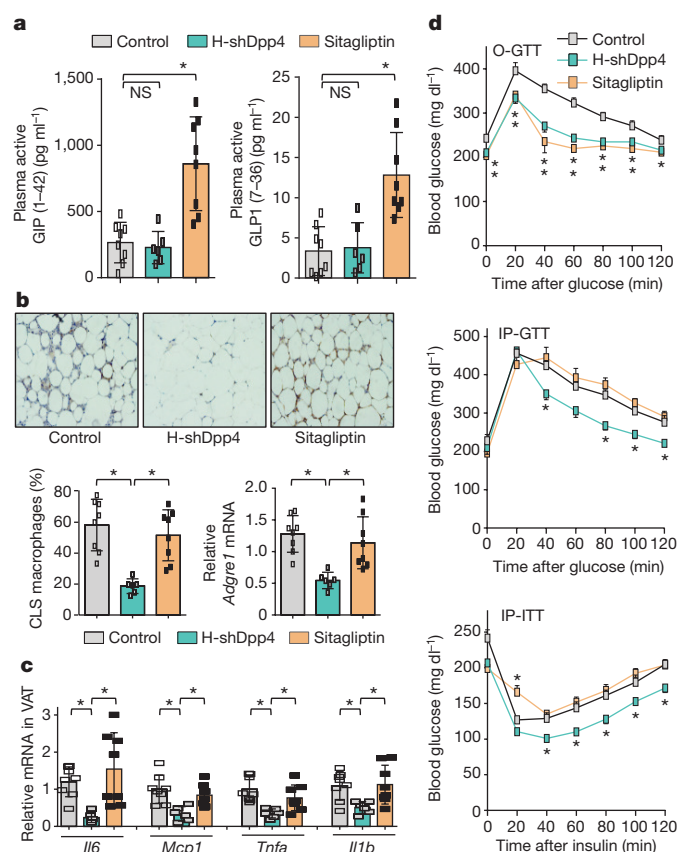


Figure 2 | Silencing of hepatocyte DPP4, but not treatment with the oral DPP4 inhibitor sitagliptin, lowers VAT inflammation and improves metabolism in DIO mice. **a–d**, Control, H-shDpp4- and sitagliptin-treated mice from those described in Extended Data Fig. 6c–h were analysed as follows. **a**, Plasma active GIP (1–42) and GLP1 (7–36). NS, not significant. **b**, CLS macrophages and *Adgre1* mRNA in VAT, with representative images of F4/80-stained VAT. **c**, *Il6*, *Mcp1*, *Tnfa* and *Il1b* mRNA in VAT. **d**, Blood glucose after oral glucose (O-GTT), intraperitoneal glucose (IP-GTT), or intraperitoneal insulin (IP-ITT). Data are mean \pm s.e.m., * P < 0.05 by one-way ANOVA.

inflammation and glucose metabolism: we directly compared treatment with the oral DPP4 inhibitor sitagliptin with silencing of hepatocyte DPP4 in DIO mice. After four weeks, neither treatment affected body weight; both treatments lowered plasma DPP4 activity to a similar degree and neither treatment affected DPP4 activity in VAT (Extended Data Fig. 6c, d). However, H-shDpp4, but not sitagliptin, decreased hepatic DPP4 protein and hepatocyte DPP4 activity (Extended Data Fig. 6e). Conversely, sitagliptin, but not H-shDpp4, increased plasma incretins (Fig. 2a), consistent with our shDpp4 data and the known action of oral DPP4 inhibitors. Notably, H-shDpp4, but not sitagliptin, suppressed VAT inflammation (Fig. 2b, c), and although both treatments lowered blood glucose and improved oral glucose tolerance, only H-shDpp4 lowered plasma insulin, improved glucose response to intraperitoneal glucose and insulin, showed evidence of increased insulin-induced p-AKT in VAT and liver and lowered plasma NEFA (Fig. 2d and Extended Data Fig. 6f–h). Consistent with these results, both sitagliptin and H-shDpp4 treatments lowered plasma DPP4 activity and blood glucose, but only H-shDpp4 decreased plasma insulin levels after 11 weeks, without a change in body weight (Extended Data Fig. 6i). Thus, DPP4 inhibition by orally administered sitagliptin and hepatocyte DPP4 silencing have fundamentally different effects on VAT inflammation and metabolism.

Next, we investigated the mechanism of DPP4-induced VAT macrophage inflammation. Although recombinant DPP4 (rDPP4) alone was unable to induce *Mcp1* or *Il6* expression in SVF, it did so when

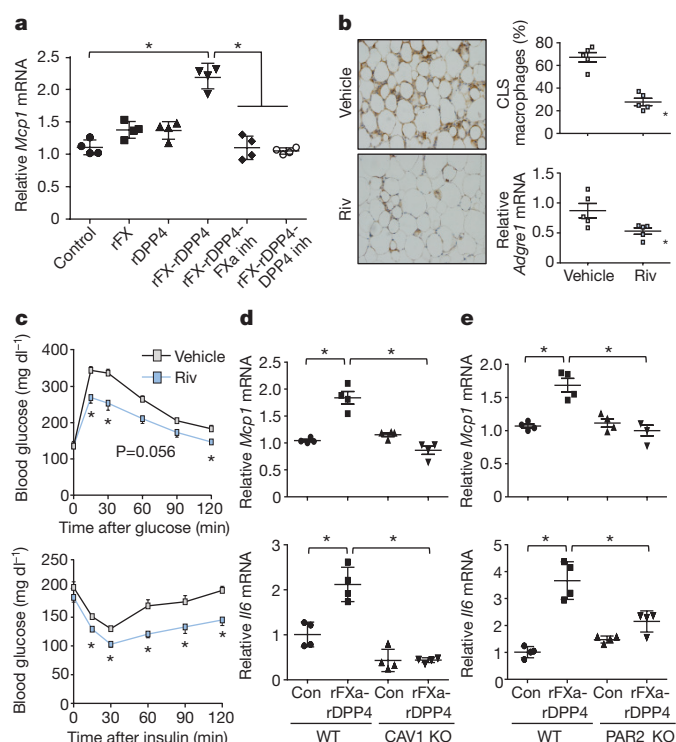


Figure 3 | DPP4 and FXa synergistically activate inflammatory signalling in macrophages. **a**, BMDMs were pretreated with or without 10 μ M FXa inhibitor rivaroxaban or 10 μ M DPP4 inhibitor KR62436, followed by incubation for 4 h with rFX or rDPP4 alone or in combination. *Mcp1* mRNA was then quantified. n = 4 technical replicates per group; mean \pm s.e.m.; * P < 0.05 by one-way ANOVA. **b**, **c**, Sixteen-week-old mice previously fed the DIO diet for 10 weeks were treated for 20 days with 2 mg kg⁻¹ oral rivaroxaban twice daily (Riv) or vehicle control and analysed as follows. **b**, CLS macrophages and *Adgre1* mRNA in VAT, with representative images of F4/80-stained VAT. **c**, Blood glucose after intraperitoneal glucose or insulin (n = 5 mice per group; mean \pm s.e.m.; * P < 0.05 by two-tailed Student's t -test). **d**, **e**, BMDMs from wild-type, CAV1 knockout (CAV1 KO) or PAR2 knockout (PAR2 KO) mice were incubated for 4 h with or without rFXa and rDPP4 and then assayed for *Mcp1* and *Il6* mRNA. n = 4 technical replicates per group; mean \pm s.e.m.; * P < 0.05 by two-way ANOVA.

added to either DIO mouse plasma that had been immunodepleted of DPP4 or plasma from DIO mice that had been treated with H-shDpp4 (Extended Data Fig. 7a, b). To determine whether an additional plasma factor was needed for DPP4 to promote SVF inflammation, plasma from DIO mice was immunodepleted of DPP4 and then fractionated using FPLC. Two fractions (F43 and F44) were able to induce *Mcp1* in SVF cells in the presence of rDPP4, but not in its absence (Extended Data Fig. 7c). Fraction F44 and the inactive fractions F42 and F46 were then analysed by LC–MS/MS to identify peptides that were more abundant in F44 versus F42 and F46 (Supplementary Table 2b). Peptides corresponding to plasma factor X met these criteria (Extended Data Fig. 8a). Factor X can trigger inflammation in endothelial cells and leukocytes^{24,25}, and we found that the ability of DIO plasma to induce *Mcp1* expression in SVF cells or macrophages was abrogated by rivaroxaban, an inhibitor of factor Xa (FXa) (Extended Data Fig. 8b). Furthermore, whereas recombinant factor X (rFX) or rDPP4 alone led to only very small increases in *Mcp1* and *Il6* expression in macrophages, combined treatment with both rFX and rDPP4 was much more potent, and this inflammatory activity was inhibited by either rivaroxaban or a DPP4 inhibitor (Fig. 3a and Extended Data Fig. 8c). These data also imply that factor X is converted to FXa by macrophages²⁶. Using recombinant FXa (rFXa) directly, we showed that rFXa and rDPP4 together were able to induce *Mcp1* and *Il6* expression in DIO mouse SVF cells but not in

lean mouse SVF cells (Extended Data Fig. 8d). Treatment of DIO mice with rivaroxaban lowered VAT inflammation and plasma NEFA, and improved metabolism, without affecting body weight (Fig. 3b, c and Extended Data Fig. 8e–j).

To understand how the combination of DPP4 and FXa induces *Mcp1* expression, we pre-treated macrophages with rFXa and then, after removing the rFXa and inhibiting residual activity with rivaroxaban, treated the cells with rDPP4. We also conducted the reverse experiment by pre-treating the cells with rDPP4 and then, after removing the rDPP4 and inhibiting residual activity with a DPP4 inhibitor, treated the cells with rFXa. Both protocols increased *Mcp1* and *Il6* mRNA expression (Extended Data Fig. 9a, b), suggesting that neither a complex between FXa and DPP4 nor a specifically ordered priming event is needed. We therefore hypothesized that each factor alone partially activates inflammatory signalling to a sub-threshold level, and that the threshold is reached only when both factors are present. Inflammation induced by rDPP4 and rFXa was markedly reduced in macrophages from mice lacking either caveolin-1 (CAV1), which mediates the inflammatory effects of DPP4 on antigen-presenting cells²⁷, or PAR2, a cell-surface receptor that mediates the inflammatory effects of FXa in macrophages²⁸ (Fig. 3d, e). Similar results were observed in VAT-derived ATMs from DIO mice, and human macrophages treated with inhibitors of CAV1 or PAR2 (Extended Data Fig. 9c, d). Consistent with the signalling-threshold hypothesis, rDPP4 alone increased phosphorylation of two reported CAV1 signalling intermediates, IRAK1 and TAK1 (Extended Data Fig. 9e), and rFXa alone increased phosphorylation of the PAR2 signalling intermediate RAF1 (Extended Data Fig. 9f), but only combined treatment with rDPP4 and rFXa was able to activate two distal inflammatory signalling molecules, ERK1/2 and NF- κ B (Extended Data Fig. 9g). Pharmacological inhibition of these signalling molecules abolished rDPP4–rFXa-induced *Mcp1* and *Il6* expression (Extended Data Fig. 9c, h–j). Together, these data suggest that DPP4 and FXa activate two separate upstream pathways that synergistically stimulate ERK1/2 and NF- κ B to induce *Mcp1* and *Il6* in ATMs.

Next, we determined whether inhibition of the above pathways in ATMs could ameliorate VAT inflammation and improve metabolism in obese mice. Genes can be silenced specifically in ATMs in *ob/ob* mice by intraperitoneal administration of short interfering RNA (siRNA) encapsulated in micrometer-sized glucan shells, referred to as glucan-encapsulated siRNA particles (GERPs)⁸. Using this method, we determined the effect of decreasing *Par2* and *Cav1* expression in ATMs on VAT inflammation and glucose metabolism in *ob/ob* mice. In a pilot experiment using FITC-labelled glucan shells in *ob/ob* mice, we verified their increased localization to macrophages in VAT (ATMs) relative to macrophages in other organs (Extended Data Fig. 10a). We then treated *ob/ob* mice with GERPs containing scrambled RNA (control), or siRNA targeting PAR2 or CAV1, over a 12-day period. Body weight was similar in all three cohorts (Extended Data Fig. 10b). PAR2 and CAV1 levels were decreased by the respective GERPs in ATMs but not in spleen (Extended Data Fig. 10c). Notably, both GERPs lowered VAT inflammation, improved glucose intolerance, increased insulin-induced p-AKT in VAT and liver and decreased plasma NEFA (Fig. 4a–d and Extended Data Fig. 10d, e).

These findings emphasize the important role of crosstalk between hepatocytes and adipose tissue in metabolic disease (Fig. 4e) and the distinct effects of silencing hepatocyte DPP4 expression versus DPP4 inhibition by oral inhibitors, which are widely used to treat type 2 diabetes. Although sitagliptin has been reported to lower adipose inflammation in obese mice²⁹, the drug was administered to young mice when they started receiving a high-fat diet, leading to an approximately 25% decrease in body weight in comparison to untreated mice. In our study, the drug was given to adult mice after obesity and VAT inflammation had developed, and it had no effect on body weight. We speculate that inhibition of circulating DPP4 activity does not block VAT inflammation in adult obese mice because some other action of oral DPP4

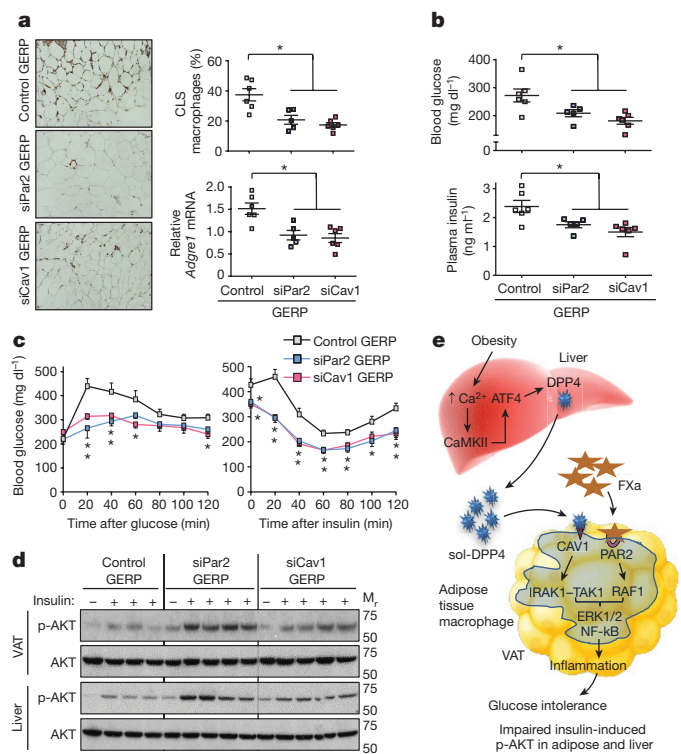


Figure 4 | Silencing PAR2 or CAV1 in ATMs in obese mice lowers VAT inflammation and improves response to insulin. **a–d**, Five-week-old, chow-fed *ob/ob* mice were injected intraperitoneally once a day for 12 days with 0.2 mg GERPs containing scrambled siRNA (control), *Par2* siRNA (siPar2) or *Cav1* siRNA (siCav1). Twenty-four hours after the last injection, the mice were analysed as follows. **a**, CLS macrophages and *Adgre1* mRNA in VAT, with representative images of F4/80-stained VAT. **b**, Blood glucose and plasma insulin 5 h after food withdrawal. **c**, Blood glucose after intraperitoneal glucose or insulin. **d**, Total AKT and p-AKT in VAT and liver after portal vein insulin injection. *n* = 5–6 mice per group; **P* < 0.05 by two-tailed Student's *t*-test for each siRNA GERP versus control GERP. For gel source data, see Supplementary Fig. 1. **e**, Summary scheme: obesity triggers a Ca²⁺–CaMKII–ATF4 pathway in hepatocytes, leading to induction of DPP4 and secretion of soluble DPP4 (sol-DPP4). Soluble DPP4 activates a CAV1–IRAK1–TAK1 pathway in ATMs, which, in combination with PAR2–RAF1 activation by FXa in ATMs, promotes ERK1/2–NF- κ B-mediated inflammation. VAT inflammation exacerbates glucose intolerance, impaired insulin signalling in adipose and liver and hyperinsulinaemia.

inhibition in that setting, perhaps related to the increase in plasma insulin³⁰, counteracts the anti-inflammatory effect of lowering plasma DPP4 activity. While future work is needed to understand the role of the hepatocyte DPP4–FXa pathway in obese, insulin-resistant humans, we have preliminary data that shows that a high percentage of obese subjects with fat inflammation and a high homeostasis model assessment (HOMA) index have inflammatory activity in their plasma that is dependent on DPP4, FXa and PAR2, whereas plasma from obese subjects without fat inflammation and insulin resistance do not show this dependency. In conclusion, the identification of this pathway provides insight into how organ crosstalk can exacerbate metabolic disarray in obesity and raises the possibility that therapeutic silencing of hepatocyte DPP4—via hepatocyte-targeted siRNA³¹, for example—may have benefits for obesity-induced insulin resistance that are not achievable through currently available oral DPP4 inhibitors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 June 2017; accepted 14 February 2018.

Published online 21 March 2018.

1. Dasgupta, S. *et al.* NF- κ B mediates lipid-induced fetuin-A expression in hepatocytes that impairs adipocyte function effecting insulin resistance. *Biochem. J.* **429**, 451–462 (2010).
2. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
3. Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
4. Hardy, O. T. *et al.* Body mass index-independent inflammation in omental adipose tissue associated with insulin resistance in morbid obesity. *Surg. Obes. Relat. Dis.* **7**, 60–67 (2011).
5. Blüher, M. Adipose tissue inflammation: a cause or consequence of obesity-related insulin resistance? *Clin. Sci. (Lond.)* **130**, 1603–1614 (2016).
6. Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. Adipose expression of tumor necrosis factor- α : direct role in obesity-linked insulin resistance. *Science* **259**, 87–91 (1993).
7. Lumeng, C. N. & Saltiel, A. R. Inflammatory links between obesity and metabolic disease. *J. Clin. Invest.* **121**, 2111–2117 (2011).
8. Aouadi, M. *et al.* Gene silencing in adipose tissue macrophages regulates whole-body metabolism in obese mice. *Proc. Natl Acad. Sci. USA* **110**, 8278–8283 (2013).
9. Ozcan, L. *et al.* Calcium signaling through CaMKII regulates hepatic glucose production in fasting and obesity. *Cell Metab.* **15**, 739–751 (2012).
10. Ozcan, L. *et al.* Activation of calcium/calmodulin-dependent protein kinase II in obesity mediates suppression of hepatic insulin signaling. *Cell Metab.* **18**, 803–815 (2013).
11. Ozcan, L. *et al.* Hepatocyte DACH1 is increased in obesity via nuclear exclusion of HDAC4 and promotes hepatic insulin resistance. *Cell Reports* **15**, 2214–2225 (2016).
12. Boonacker, E. & Van Noorden, C. J. The multifunctional or moonlighting protein CD26/DPP4. *Eur. J. Cell Biol.* **82**, 53–73 (2003).
13. Kirino, Y., Sei, M., Kawazoe, K., Minakuchi, K. & Sato, Y. Plasma dipeptidyl peptidase 4 activity correlates with body mass index and the plasma adiponectin concentration in healthy young people. *Endocr. J.* **59**, 949–953 (2012).
14. Lamers, D. *et al.* Dipeptidyl peptidase 4 is a novel adipokine potentially linking obesity to the metabolic syndrome. *Diabetes* **60**, 1917–1925 (2011).
15. Drucker, D. J. & Nauck, M. A. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet* **368**, 1696–1705 (2006).
16. Guilherme, A., Virbasius, J. V., Puri, V. & Czech, M. P. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat. Rev. Mol. Cell Biol.* **9**, 367–377 (2008).
17. Ristic, S., Byiers, S., Foley, J. & Holmes, D. Improved glycaemic control with dipeptidyl peptidase-4 inhibition in patients with type 2 diabetes: vildagliptin (LAF237) dose response. *Diabetes Obes. Metab.* **7**, 692–698 (2005).
18. Aschner, P. *et al.* Effect of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy on glycaemic control in patients with type 2 diabetes. *Diabetes Care* **29**, 2632–2637 (2006).
19. Raz, I. *et al.* Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy in patients with type 2 diabetes mellitus. *Diabetologia* **49**, 2564–2571 (2006).
20. Kutoh, E. & Ukai, Y. Alogliptin as an initial therapy in patients with newly diagnosed, drug naïve type 2 diabetes: a randomized, control trial. *Endocrine* **41**, 435–441 (2012).
21. Kadowaki, T. & Kondo, K. Efficacy, safety and dose-response relationship of teneligliptin, a dipeptidyl peptidase-4 inhibitor, in Japanese patients with type 2 diabetes mellitus. *Diabetes Obes. Metab.* **15**, 810–818 (2013).
22. Jung, C. H. *et al.* A randomized, double-blind, placebo-controlled, phase II clinical trial to investigate the efficacy and safety of oral DA-1229 in patients with type 2 diabetes mellitus who have inadequate glycaemic control with diet and exercise. *Diabetes Metab. Res. Rev.* **31**, 295–306 (2015).
23. Mulvihill, E. E. *et al.* Cellular sites and mechanisms linking reduction of dipeptidyl peptidase-4 activity to control of incretin hormone action and glucose homeostasis. *Cell Metab.* **25**, 152–165 (2017).
24. Senden, N. H. *et al.* Factor Xa induces cytokine production and expression of adhesion molecules by human umbilical vein endothelial cells. *J. Immunol.* **161**, 4318–4324 (1998).
25. Busch, G. *et al.* Coagulation factor Xa stimulates interleukin-8 release in endothelial cells and mononuclear leukocytes: implications in acute myocardial infarction. *Arterioscler. Thromb. Vasc. Biol.* **25**, 461–466 (2005).
26. McGee, M. P., Wallin, R., Wheeler, F. B. & Rothberger, H. Initiation of the extrinsic pathway of coagulation by human and rabbit alveolar macrophages: a kinetic study. *Blood* **74**, 1583–1590 (1989).
27. Ohnuma, K. *et al.* CD26 mediates dissociation of Tollip and IRAK-1 from caveolin-1 and induces upregulation of CD86 on antigen-presenting cells. *Mol. Cell. Biol.* **25**, 7743–7757 (2005).
28. Zuo, P. *et al.* Factor Xa induces pro-inflammatory cytokine expression in RAW 264.7 macrophages via protease-activated receptor-2 activation. *Am. J. Transl. Res.* **7**, 2326–2334 (2015).
29. Dobrian, A. D. *et al.* Dipeptidyl peptidase IV inhibitor sitagliptin reduces local inflammation in adipose tissue and in pancreatic islets of obese mice. *Am. J. Physiol. Endocrinol. Metab.* **300**, E410–E421 (2011).
30. Pedersen, D. J. *et al.* A major role of insulin in promoting obesity-associated adipose tissue inflammation. *Mol. Metab.* **4**, 507–518 (2015).
31. Fitzgerald, K. *et al.* A highly durable RNAi therapeutic inhibitor of PCSK9. *N. Engl. J. Med.* **376**, 41–51 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. S. Katz for assistance with FPLC; R. Kaufman for adeno-ATF4; C. Adams and S. Bullard for *Atf4*^{fl/fl} mice; and A. Ferrante, S. Ramakrishnan, J. Weitz and T. McGraw for discussions. E.C. was supported by NIH grant 5P30CA013696-42. I.T. was funded by grants from the NIH (HL087123 and HL075662) and by a grant from the Merck Investigator Studies Program. L.O. was funded by the NIH grant DK106045 and a grant from the Columbia University Diabetes Research Center (P30 DK063608). Y.S., S.M.N. and M.P.C. were funded by NIH grant DK103047. M.B. was funded by the Deutsche Forschungsgemeinschaft grant SFB1052.

Author Contributions D.S.G., L.O. and I.T. designed the study, analysed data and wrote the manuscript. D.S.G., L.O. and Z.Z. conducted the experiments. S.M.N., Y.S. and M.P.C. made the glucan-encapsulated siRNA particles (GERPs) and helped design these experiments and analyse the data. E.C. conducted the LC-MS/MS studies and assisted with data analysis. M.B. helped with interpretation of data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.O. (lo2192@columbia.edu) or I.T. (iat1@columbia.edu).

Reviewer Information *Nature* thanks P. Scherer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Reagents and antibodies. Recombinant mouse and human DPP4 (catalogue no. 954-SE and 9168-SE, respectively) and anti-DPP4 antibody (AF954) were from R&D Systems. The PAR2 inhibitor GB83 was from Axon Medchem (1622). Recombinant FX (233282) and FXa (233526), and the MEK inhibitor PD98059 (513000) were from Millipore. The FXa inhibitor rivaroxaban (S3002) was from Selleckchem. IRAK1-4 inhibitor I (A3505) and sitagliptin phosphate monohydrate (A4036) were from ApexBio. Tri reagent (T9424), sodium palmitate (P9767), collagenase (C2139), liberase (5401020001), insulin (I9278), the DPP4 inhibitor KR-62436 hydrate (K4264), the caveolin-1 inhibitor daidzein (D7802), PS-1145, IKK inhibitor (P6624), the TAK1 MAPKKK inhibitor 5Z-7-oxozeaenol (O9890), anti-phospho-Thr209-IRAK1 (SAB4504246), and anti- β -actin antibody (A3854) were from Sigma. Anti-phospho-S473-AKT (4060), anti-AKT (4691), anti-phospho-Thr202/Tyr204-ERK1/2 (8544), anti-ERK (9194), anti-phospho-S536-NF- κ B p65 (3033), anti-NF- κ B p65 (8242), anti-phospho-S412-TAK1 (9339), anti-TAK1 (5206), anti-phospho-S338-cRAF1 (9427), anti-cRAF1 (53745) and anti-caveolin-1 (3267) antibodies were from Cell Signaling. Anti-PAR2 antibody (817201) was from Biologend. The DIO high-fat diet (60% kcal from fat) (D12492) was from Research Diets. Adenoviral *Atf4* (adeno-*Atf4*) was a gift from R. J. Kaufman (Sanford-Burnham Medical Research Institute), adenoviral *Dpp4* (adeno-*Dpp4*) was purchased from Vector Biolabs, and adeno-associated virus 8 (AAV8) containing either hepatocyte-specific TBG-Cre recombinase (AAV8-TBG-cre) or the control vector (AAV8-TBG-lacZ) were purchased from the Penn Vector Core. All adenoviruses were amplified by Viraquest. The AAV8-H1-shRNA construct targeting murine DPP4 was made by annealing complementary oligonucleotides and then ligating them into the pAAV-RSV-GFPH1 vector, as described previously³². The resultant constructs were amplified by Salk Institute Gene Transfer, Targeting, and Therapeutics Core. siRNA sequences against murine *Par2* were purchased from GE Dharmacon and siRNA sequences against mouse *Cav1* were purchased from Integrated DNA Technologies.

Mouse experiments. *Camk2g^{fl/fl}* mice were generated as previously described¹⁰ and crossed onto the C57BL/6J background. *Atf4^{fl/fl}* mice were generated as previously described³³. Male *Camk2g^{fl/fl}* or *Atf4^{fl/fl}* mice were fed a high-fat, high-calorie diet (HFD, 60% kcal from fat) for 13 weeks starting at three weeks of age and were maintained on a 12-h light-dark cycle. Fifteen-week-old high-fat-diet-fed obese mice (DIO) that had been fed with HFD for nine weeks, four-week old chow-fed *ob/ob* mice, or sixteen-week-old chow-fed wild-type C57BL/6J mice were purchased from Jackson Laboratories and maintained for one week in the animal facility before commencement of experiments. Sixteen-week-old DIO *Camk2g^{fl/fl}* mice that had been fed with HFD for 13 weeks were treated with adeno-associated viruses (AAV) for expression of either hepatocyte-specific TBG-cre³⁴ or TBG-lacZ (the control vector) for three weeks to obtain H-CaMKII(KO) and wild-type mice. To obtain ATF4-restored H-CaMKII(KO) or DPP4-restored H-CaMKII(KO) mice, TBG-cre-treated *Camk2g^{fl/fl}* mice received either adeno-*Atf4* (H-CaMKII(KO) plus adeno-*Atf4*) or adeno-DPP4 (H-CaMKII(KO) plus adeno-*Dpp4*), while the remaining mice received adeno-lacZ. Similarly, sixteen-week-old *Atf4^{fl/fl}* DIO mice that had been fed with HFD for 13 weeks were treated with AAV8-TBG-cre or AAV8-TBG-lacZ to obtain hepatocyte-ATF4 KO (H-ATF4(KO)) and wild-type mice. Sixteen-week-old wild-type DIO mice, four-week old *ob/ob* and sixteen-week-old wild-type C57BL/6J mice were treated with AAV8-H1-shDpp4 or AAV8-H1-control to obtain hepatocyte-specific DPP4 silencing (H-shDpp4) or control mice, respectively. In all mouse experiments, recombinant adenovirus (1×10^9 plaque-forming units per mouse) or adeno-associated virus (1×10^{11} genome copies per mouse) was delivered by tail vein injection, and experiments were commenced after 7–28 days or, for the experiment in Extended Data Fig. 6i, after 11 weeks. DIO mice were treated with oral rivaroxaban (2 mg/kg) twice daily for 20 days. DIO mice were treated for four or seven weeks with sitagliptin by adding it to the drinking water at 0.3 mg/ml, which results in a dose of ~30–45 mg/kg/day. Mouse plasma samples were collected from lean or obese mice after 4–5 h of food withdrawal, with free access to water. Blood glucose was measured after 4–5 h of food withdrawal in mice using a glucose meter (One Touch Ultra, Lifescan). Plasma insulin levels were measured using the Ultrasensitive Mouse Insulin ELISA Kit (Crystal Chem, 90080) and plasma MCP1, IL6 and TNF α levels were measured using mouse ELISA kits from RayBiotech (ELM-MCP1-1; ELM-IL6-1; ELM-TNF α -1). For gastric inhibitory polypeptide (GIP) and glucagon-like peptide-1 (GLP-1) measurements, mice were fasted overnight and then fed for 4–5 h before collecting plasma samples into a tube containing DPP4 inhibitor (5 μ M). Plasma active GIP was measured using Sandwich mouse active GIP (1–42) ELISA Kit (Crystal Chem, 81511) and GLP-1 was assayed in plasma samples using mouse active GLP-1 (7–36) ELISA kit (Eagle Biosciences, GP121-K01). Plasma non-esterified fatty acids (NEFA) were assayed using an enzymatic kit from Wako Diagnostics (HR Series NEFA-HR(2)-Colour Reagent A, 999-34691; Solvent A, 995-34791;

Colour Reagent B, 991-34891; Solvent B, 993-35191; and NEFA standard solution, 276-76491). Intraperitoneal-glucose tolerance tests were performed by intraperitoneal injection of glucose (1 g/kg body weight for DIO, 0.5 g/kg body weight for *ob/ob* and 2 g/kg body weight for lean mice) following an overnight fast. Oral glucose tolerance tests were conducted by delivering a 2-g/kg glucose bolus orally via gavage to DIO mice after a 5-h fast. Insulin tolerance tests were performed in mice that were fasted for 4–5 h, by intraperitoneal insulin injection (0.6 IU/kg body weight for DIO, 2 IU/kg body weight for *ob/ob* and 0.5 IU/kg body weight for lean mice) and by assaying blood glucose at various time points. Previous studies and pilot experiments formed the basis of power calculations for the various studies. Depending on the experiment, calculations indicated that 3–12 mice per group would enable the testing of our hypotheses based on an expected 25–30% coefficient of variation and an 80% chance of detecting a 33% difference in the key specified endpoints ($P < 0.05$). For all experiments, male mice of the same age and similar weight were randomly assigned to experimental and control groups. On occasion, we analysed a subset of mice for a particular parameter, and the subset was chosen randomly from a full cohort. Pre-specified exclusion criteria were weight loss 10% of initial body weight or signs of illness or injury requiring euthanasia. According to these pre-specified criteria, the maximum number of mice removed before analysis was three, but was more typically between zero and two. Animal studies were conducted in accordance with the Columbia University Animal Research Committee. Moreover, because switching mice from group caging to single caging, improper handling and placing mice in a restrainer affect stress-sensitive metabolic parameters including plasma DPP4 levels and adipose tissue inflammation (D.S.G., L.O. and I.T., unpublished observations), all experimental mice were housed in groups to avoid stress from isolation and handled with extra care to avoid stress during blood glucose measurement and blood collection.

F4/80 immunostaining of VAT. VAT samples were fixed in 10% (v/v) formalin solution for 24 h, followed by incubation with 70% ethanol (v/v) for 24 h and then embedding in paraffin. Six-micrometre sections were mounted on charged glass slides, incubated at 60 °C for 20 min, and deparaffinized using three washes with xylene. Antigen retrieval was carried out by treating the sections with proteinase K (1:1,000) (Invitrogen, AM2546) at 37 °C for 25 min. The sections were then rinsed and blocked using 5% serum, followed by incubation overnight at 4 °C with anti-F4/80 biotin-conjugated primary antibody (1:100) (Miltenyi Biotec, 130-101-893). The sections were then incubated with streptavidin-HRP secondary antibody (1:200) (BD Pharmingen, 554066) and then developed in chromagen substrate 3,3'-diaminobenzidine (Cell Signaling, 8059) and counterstained with haematoxylin. For each mouse, 5–10 sections were analysed by microscopic imaging at 20 \times magnification. The total number of nuclei and the number of nuclei of F4/80-immunostained cells surrounding adipocytes in CLS were counted for each field. The data are expressed as mean percentage of CLS macrophages per total adipose tissue nuclei.

Portal vein insulin infusion and protein extraction from tissues. After 4–5 h of food withdrawal, mice were anaesthetized and insulin (0.6 IU/kg body weight for DIO, 2 IU/kg body weight for *ob/ob* and 0.5 IU/kg body weight for lean mice) or PBS was injected through the portal vein. Three minutes after injection, liver and adipose tissue were removed, frozen in liquid nitrogen, and kept at –80 °C until processing. For protein extraction, tissue samples were cut into small pieces and transferred to tubes containing ice-cold RIPA buffer supplemented with Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific, 78440). Tissue segments were then homogenized on ice, lysates were centrifuged, and the supernatant fractions were used for immunoblot analysis.

Primary hepatocyte experiments. Wild-type C57BL/6J mice (8–10 weeks old) were injected with either control vector or AAV8-shDpp4 (H-shDpp4) via tail vein injection. Ten days after injection, primary mouse hepatocytes were isolated as previously described³⁵. Hepatocytes were incubated with either BSA control or palmitate (50 μ M) for 10 h, with the last 5 h in serum-free medium. The cells were then treated with 100 nM insulin or vehicle control for 5 min and immediately frozen in liquid nitrogen and kept at –80 °C until further processing. Cells were lysed using 2 \times Laemmli buffer supplemented with Halt Protease and Phosphatase Inhibitor Cocktail and lysates were used for immunoblotting.

Preparation of stromal vascular fraction cells and isolation of adipose tissue macrophages. SVF from lean or obese mice was prepared as previously described³⁶. In brief, VAT was isolated in ice-cold PBS containing liberase (50 μ g/ml) and minced into small segments. VAT was digested at 37 °C for 1 h with intermittent mixing. After digestion, the solution was centrifuged, buoyant adipocytes were removed, and the cell pellet was retrieved as SVF. SVF cells were then cultured in medium containing 10% (vol/vol) FBS or plasma from lean or obese mice. For isolation of ATMs, anti-F4/80 biotin-conjugated primary antibody (Miltenyi Biotec, 130-101-893) was added at a 1:10 dilution to 4×10^6 SVF cells in 200 μ l of PBS containing 5% of FBS (PBS-FBS), followed by incubation at 4 °C for 20 min.

The SVF cells were pelleted by centrifugation, washed once with 1 ml of buffer, and resuspended in 200 μ l of PBS–FBS. The cells were then incubated with streptavidin-conjugated magnetic microbeads (1:10) (Miltenyi Biotec, 130-048-101) at 4°C for 20 min. The cells were rinsed once with PBS–FBS, and ATMs were isolated using magnetic separation columns (Miltenyi Biotec, 130-042-201) and cultured in medium containing 10% (vol/vol) FBS.

Immunoblotting. Proteins from tissue or cell lysates were resolved on 4–20% Tris–glycine gradient gels and then transferred to nitrocellulose membranes. The membranes were blocked using either 5% non-fat milk or 5% BSA followed by incubation with primary antibodies overnight. Blots were then washed thoroughly and probed with HRP-conjugated secondary antibodies for 1 h. Protein bands were detected with Supersignal West Pico enhanced chemiluminescent solution (Thermo Scientific, Cat. No. 34080) and analysed by ImageJ.

Quantitative reverse transcription–PCR. Total RNA was extracted from tissues or primary cell cultures using Tri-reagent (Sigma, T9424). cDNA was synthesized from 2 μ g total RNA using oligo (dT) and Superscript II (Invitrogen, AM5730G and 18064014). Real time quantitative (q)PCR was performed with a 7500 Real-Time PCR System (Applied Biosystems) using SYBR Green master mix (Qiagen, 204145) and gene-specific primers. The sequences of gene-specific primers used for the qPCR assays are described in Supplementary Table 1.

DPP-4 immunodepletion from plasma. DPP-4 was immunodepleted from obese mouse plasma using bead-immobilized anti-DPP-4 antibody (R&D Systems, AF954) as previously described³⁷. In brief, 10 μ g of anti-DPP-4 antibody was conjugated with 1.5 mg of Protein-G Dynabeads (Invitrogen, 10004D) in 200 μ l of binding buffer. After immobilization of anti-DPP-4 antibody to beads, 40 μ l of DIO mouse plasma and 60 μ l of PBS were mixed with 40 μ l of antibody-coated bead suspension. The resultant mixture was incubated for 2.5 h at 4°C on a rotating wheel. After 2.5 h, Dynabeads were separated with a magnet and discarded. This was followed by one more round of immunocapture of DPP4 and discarding of beads. DPP-4 proteolytic activity assay was performed on immunodepleted plasma to confirm DPP-4 immunodepletion. For some experiments, DIO mouse plasma was incubated with anti-DPP-4-conjugated or control IgG-conjugated beads, followed by elution of the bound material using 50 mM glycine buffer, pH 2.8.

Separation of plasma proteins using FPLC. Lean or obese mouse plasma (0.2 ml) was diluted 1:1 with PBS and injected onto a Superdex 200 Increase 10/300GL FPLC column (GE Healthcare, 28990944). Plasma proteins were eluted from the column at a constant flow of 0.3 ml/min with PBS (pH 7.4). The effluent was collected in 0.25-ml fractions. The collected fractions were frozen in dry ice and kept at –80°C until further processing.

In-gel protein digestion. Inflammatory obese mouse plasma, DPP4-depleted obese mouse plasma fractions or DPP4 immunoprecipitated from obese mouse plasma (IP-DPP4) were resolved on a 4–20% Tris–glycine gradient gel and then stained with GelCode Blue Safe Coomassie stain (Thermo Scientific, 24594) for 2 h followed by de-staining with double distilled water overnight. The resolved proteins were excised and processed at Herbert Irving Comprehensive Cancer Center Proteomics Shared Resource, Columbia University. The excised gel pieces were rehydrated and digested in 80 μ l of 12.5 ng/ μ l Trypsin Gold in 50 mM ammonium bicarbonate at 37°C overnight. After digestion was complete, condensed evaporated water was collected from the tube walls by brief centrifugation using a benchtop microcentrifuge (Eppendorf). The gel pieces and digestion reaction were mixed with 50 μ l 2.5% trifluoroacetic acid (TFA) and rigorously mixed for 15 min. The solution with extracted peptides was transferred into a fresh tube. The remaining peptides were extracted with 80 μ l 70% acetonitrile–5% TFA mixture using rigorous mixing for 15 min. The extracts were pooled and dried to completion (1.5–2 h) in a SpeedVac. The dried peptides were reconstituted in 30 μ l 0.1% TFA by mixing for 5 min and stored in ice or at –20°C before analysis.

LC–MS/MS analysis. The concentrated peptide mix was reconstituted in a solution containing 2% acetonitrile and 2% formic acid for mass spectrometry analysis. Peptides were eluted from the column using a Dionex Ultimate 3000 Nano liquid chromatography system with a 10-min gradient from 2% buffer B to 35% buffer B (100% acetonitrile, 0.1% formic acid). The gradient was switched from 35% to 85% buffer B over 1 min and held constant for 2 min. Finally, the gradient was changed from 85% buffer B to 98% buffer A (100% water, 0.1% formic acid) over 1 min, and then held constant at 98% buffer A for a further 5 min. The application of a 2.0-kV distal voltage electrosprayed the eluting peptides directly into the Thermo Fusion Tribrid mass spectrometer equipped with an EASY-Spray source (Thermo Scientific). Mass spectrometer-scanning functions and HPLC gradients were controlled by the Xcalibur data system (Thermo Finnigan).

Database search and interpretation of MS/MS data. Tandem mass spectra from raw files were searched against a mouse protein database using the Proteome Discoverer 1.4 software (Thermo Finnigan). The Proteome Discoverer application extracts relevant MS/MS spectra from the .raw file and determines the precursor charge state and the quality of the fragmentation spectrum. The Proteome

Discoverer probability-based scoring system rates the relevance of the best matches found by the SEQUEST algorithm. The mouse protein database was downloaded as FASTA-formatted sequences from the Uniprot protein database (database released in September, 2014). The peptide mass search tolerance was set to 10 p.p.m. A minimum sequence length of seven amino acids residues was required. Only fully tryptic peptides were considered. To calculate confidence levels and false discovery rates (FDR), Proteome Discoverer generates a decoy database containing reverse sequences of the non-decoy protein database and performs the search against this concatenated database (non-decoy + decoy). Scaffold (Proteome Software) was used to visualize searched results. The discriminant score was set at less than 1% FDR, determined on the basis of the number of accepted decoy database peptides, to generate protein lists for this study. Spectral counts were used for estimation of relative protein abundance between samples.

Mouse liver nuclei preparation and chromatin immunoprecipitation assays. Mouse liver tissues were homogenized using a Dounce homogenizer (Wheaton) with the loose pestle in 1:10 (w:v) of ice-cold NP-40 lysis buffer supplemented with protease inhibitor cocktail. Liberation of nuclei was monitored by DAPI staining and fluorescence microscopy. To purify intact nuclei, lysates were then layered over a sucrose gradient consisting of 1 M sucrose overlaid with 0.68 M sucrose and then centrifuged at 4000 r.p.m. for 20 min at 4°C. Following a washing step, nuclear pellets were cross-linked with 1% fresh formaldehyde in PBS for 10 min at room temperature. Cross-linking was terminated by addition of 200 mM Tris-HCl (pH 9.4) and 1 mM DTT for 10 min followed by centrifugation at 2,500 r.p.m. for 10 min at 4°C. Nuclear pellets were suspended in SDS lysis buffer containing protease inhibitors, incubated for 10 min on ice, and DNA was sheared in a cold water bath in a focused-ultrasonicator (Covaris, S2) to obtain DNA fragments with an average size of 500 bp. Fragmented chromatin was pre-cleaned by incubating with normal rabbit IgG (Santa Cruz, sc-2025) for 1 h at 4°C, followed by 1 h of incubation with 50 μ l protein G magnetic beads (Pierce) at 4°C with rotation. A rabbit anti-ATF4 antibody (Cell Signaling, 11815) to pull down ATF4-binding complexes, and a control rabbit anti-haemagglutinin antibody (Santa Cruz) were used to pull down non-specific binding complexes. Immunoprecipitated chromatin fragments were reverse cross-linked, digested by proteinase K, and purified using QIAquick PCR Purification Kit (Qiagen, 28104). The presence of ATF4 in the *Dpp4* exonic region was quantified by qPCR and expressed relative to the input genomic DNA. The sequences of primers used for the ChIP–qPCR assays are described in Supplementary Table 2.

Macrophage assays. ATMs, peritoneal macrophages or BMDMs were incubated for 4 h in medium containing 10% plasma from lean or obese mice and then assayed for *Mcp1* or *Il6* mRNA using quantitative reverse transcription–PCR (RT–qPCR). ATMs, bone marrow-derived macrophages or human macrophages were incubated for 4 h with 0.25 U/ml rFXa, 175 ng/ml rDPP4, or both proteins together. The cells were then assayed for *Mcp1* or *Il6* mRNA using RT–qPCR.

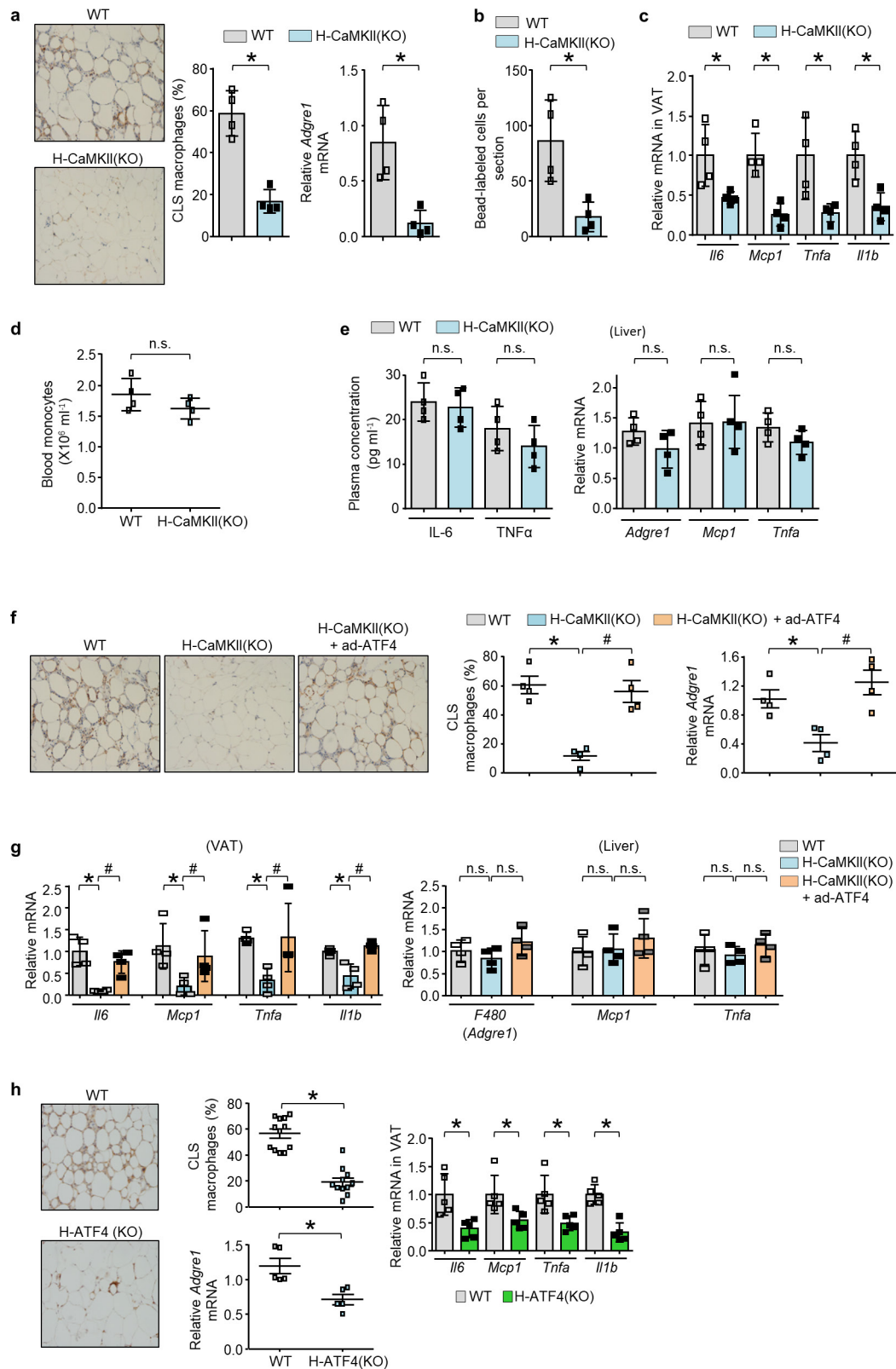
Glucan-encapsulated siRNA particle (GERP) study in *ob/ob* mice. FITC-labelled glucan shells (GS) containing non-targeting (control) siRNA or siRNAs against *Par2* or *Cav1* were prepared and loaded as previously described^{38,39}. In brief, 2.4 mg of FITC-labelled glucan shells–lectin, 7.2 nmol of siRNA, and 120 nmol of endoporter were combined to form FITC-GERPs. Five-week old *ob/ob* mice received 12 daily doses of FITC-GERPs by intraperitoneal injection, which contained 0.2 mg FITC-GS, 0.6 nmol siRNA and 10 nmol endoporter. Glucose tolerance and insulin tolerance tests were performed on days seven and ten of GERP treatment, respectively. Mice were euthanized 24 h after the last injection.

Statistical analysis. All results are presented as mean \pm s.e.m. For experiments with two groups, *P* values were calculated using the two-tailed Student's *t*-test for normally distributed data and the Mann–Whitney rank sum test for non-normally distributed data. For experiments with more than two groups, *P* values were calculated using one-way or two-way ANOVA for normally distributed data and the Kruskal–Wallis by ranks test for non-normally distributed data, with Student–Newman–Keuls or Dunn's post hoc test. Error bars that appear to be absent from certain graph symbols signify s.e.m. values smaller than the graph symbols.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

32. Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382–386 (2014).
33. Ebert, S. M. *et al.* Stress-induced skeletal muscle Gadd45a expression reprograms myonuclei and causes muscle atrophy. *J. Biol. Chem.* **287**, 27290–27301 (2012).
34. Mu, X. *et al.* Hepatocellular carcinoma originates from hepatocytes and not from the progenitor/biliary compartment. *J. Clin. Invest.* **125**, 3891–3903 (2015).
35. Shimano, H. *et al.* Overproduction of cholesterol and fatty acids causes massive liver enlargement in transgenic mice expressing truncated SREBP-1a. *J. Clin. Invest.* **98**, 1575–1584 (1996).

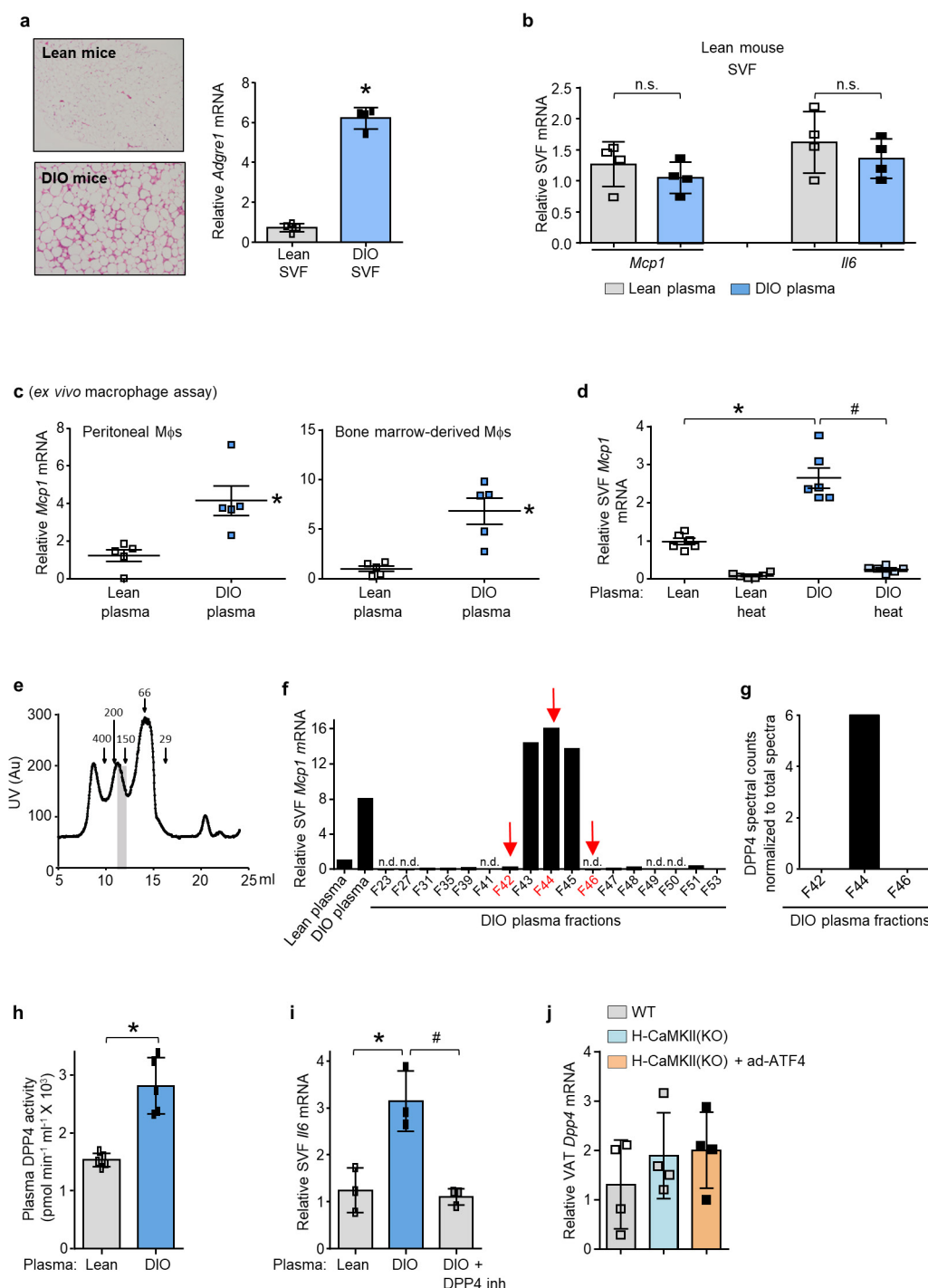
36. Orr, J. S., Kennedy, A. J. & Hasty, A. H. Isolation of adipose tissue immune cells. *J. Vis. Exp.* **22**, e50707 (2013).
37. Nazarian, A. *et al.* Inhibition of circulating dipeptidyl peptidase 4 activity in patients with metastatic prostate cancer. *Mol. Cell. Proteomics* **13**, 3082–3096 (2014).
38. Aouadi, M. *et al.* Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation. *Nature* **458**, 1180–1184 (2009).
39. Cohen, J. L. *et al.* Peptide- and amine-modified glucan particles for the delivery of therapeutic siRNA. *Mol. Pharm.* **13**, 964–978 (2016).



Extended Data Figure 1 | See next page for caption.

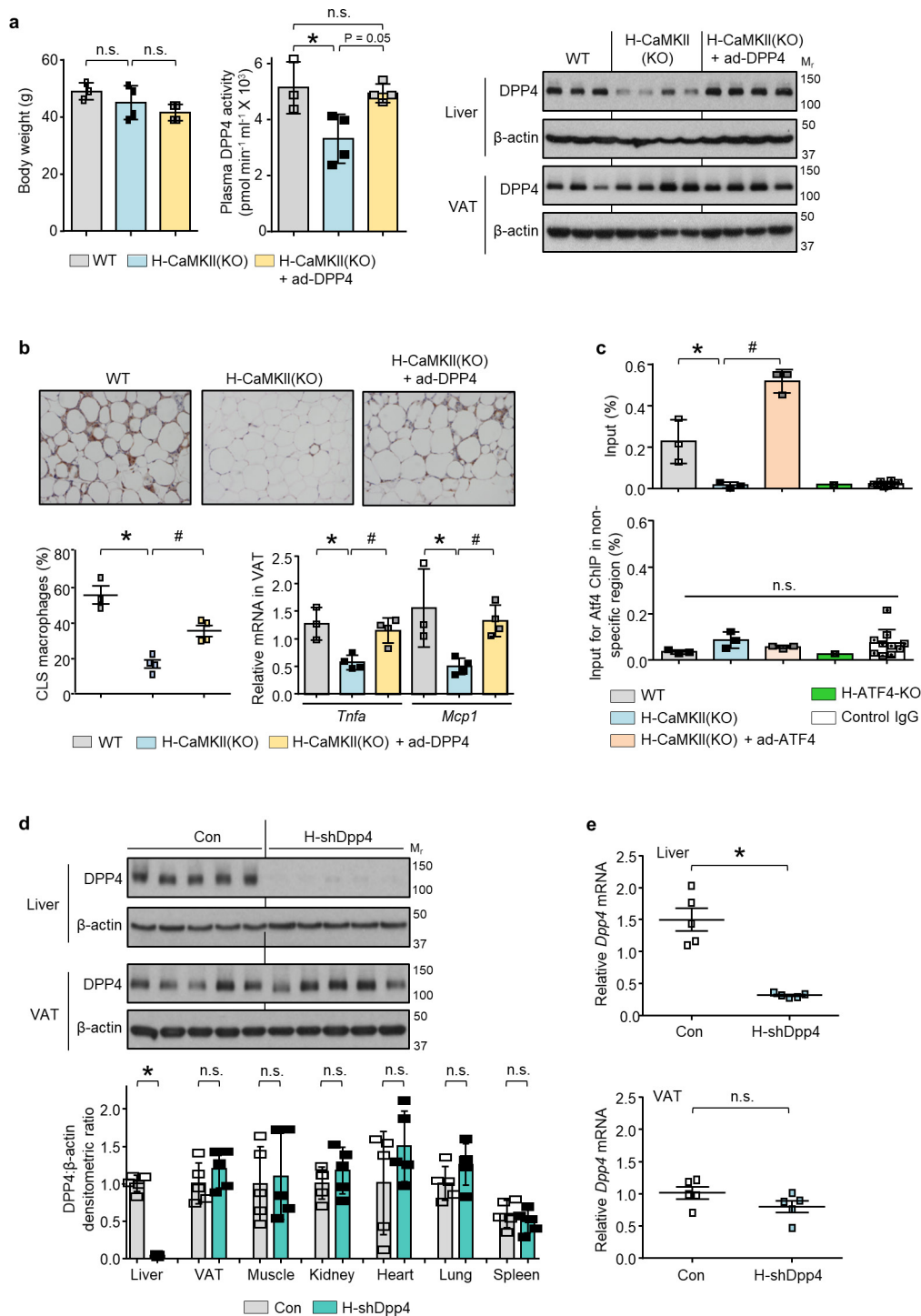
Extended Data Figure 1 | Hepatocyte-specific deletion of CaMKII or ATF4 in DIO mice lowers VAT inflammation. **a–c**, Sixteen-week-old *Camk2g^{fl/fl}* mice that had been previously fed the DIO diet for 13 weeks were injected intravenously with AAV8-TBG-*cre* (H-CaMKII(KO)) or AAV8-TBG-*lacZ* (wild-type, WT). Mice were analysed after three additional weeks on the DIO diet. **a**, Representative images of VAT immunostained for F4/80, with quantification of crown-like-structure (CLS) macrophages, and expression of *Adgre1* mRNA, which encodes F4/80. **b**, As in **a**, except that the mice were injected with fluorescent beads using a procedure that labels circulating Ly6c^{hi} monocytes, and then bead-labelled cells were assayed in VAT sections. **c**, mRNAs for *Il6*, *Mcp1*, *Tnf* and *Il1b* in VAT. **d**, Blood monocyte count. **e**, Plasma IL6 and TNF α measured by ELISA, and quantification of *Adgre1*, *Mcp1* and *Tnfa* mRNA in liver. In **a–e**, $n = 4$ mice per group; mean \pm s.e.m.; * $P < 0.05$; n.s., not significant by two-tailed Student's *t*-test. **f, g**, Mice similar to those in **a–c**, and from a third group in which *Camk2g^{fl/fl}* mice were injected

intravenously with adeno-*Atf4* and AAV8-TBG-*cre* (H-CaMKII(KO) + adeno-*Atf4*). **f**, CLS macrophages and *Adgre1* mRNA in VAT were quantified, with representative images of F4/80-stained VAT. **g**, *Il6*, *Mcp1*, *Tnfa* and *Il1b* mRNA in VAT and *Adgre1*, *Mcp1* and *Tnfa* mRNA in liver were quantified. Note that the first two groups of mice received adeno-*lacZ* instead of adeno-*Atf4*. $n = 4$ mice per group; mean \pm s.e.m.; * $P < 0.05$ by one-way ANOVA; n.s., not significant. **h**, AAV8-TBG-*cre* (H-ATF4(KO)) or AAV8-TBG-*lacZ* (wild-type) was injected intravenously into 16-week-old *Aft4^{fl/fl}* mice previously fed the DIO diet for 13 weeks. After three further weeks on the DIO diet, VAT from these mice was immunostained for F4/80 to identify macrophages, the percentage of macrophages in CLS was quantified, and the VAT was assayed for *Adgre1* and the indicated inflammatory mRNAs. Twelve wild-type and 11 H-ATF4(KO) mice was analysed for CLS macrophages, and a randomly selected subset of five wild-type and five H-ATF4(KO) mice was analysed for the VAT mRNAs. Mean \pm s.e.m., * $P < 0.05$ by two-tailed Student's *t*-test.



Extended Data Figure 2 | DPP4 in the plasma of DIO mice induces *Mcp1* and *Il6* in SVF cells from the VAT of obese mice and in macrophages. **a**, Representative images of haematoxylin and eosin-stained VAT sections from lean and DIO mice and *Adgre1* mRNA levels in SVF from lean and DIO mice. *n* = 4 mice per group; mean \pm s.e.m.; **P* < 0.05. **b**, SVF cells from the VAT of lean mice were incubated for 4 h in medium containing 10% (v/v) plasma from lean or DIO mice and then assayed for *Mcp1* and *Il6* mRNA. *n* = 4 mice per group; mean \pm s.e.m.; **P* < 0.05; n.s., not significant. **c**, *Mcp1* mRNA was assayed in mouse peritoneal macrophages (M ϕ s) or bone marrow-derived macrophages that were incubated for 4 h with medium containing 10% (v/v) plasma from lean or DIO mice. *n* = 5 technical replicates per group; mean \pm s.e.m.; **P* < 0.05. **d**, *Mcp1* mRNA levels in SVF cells that were incubated for 4 h with medium containing 10% (v/v) control or heat-inactivated (heat) plasma from the indicated groups of mice. *n* = 6 mice per group; mean \pm s.e.m.; **P* < 0.05. **e**, UV protein chromatogram obtained after

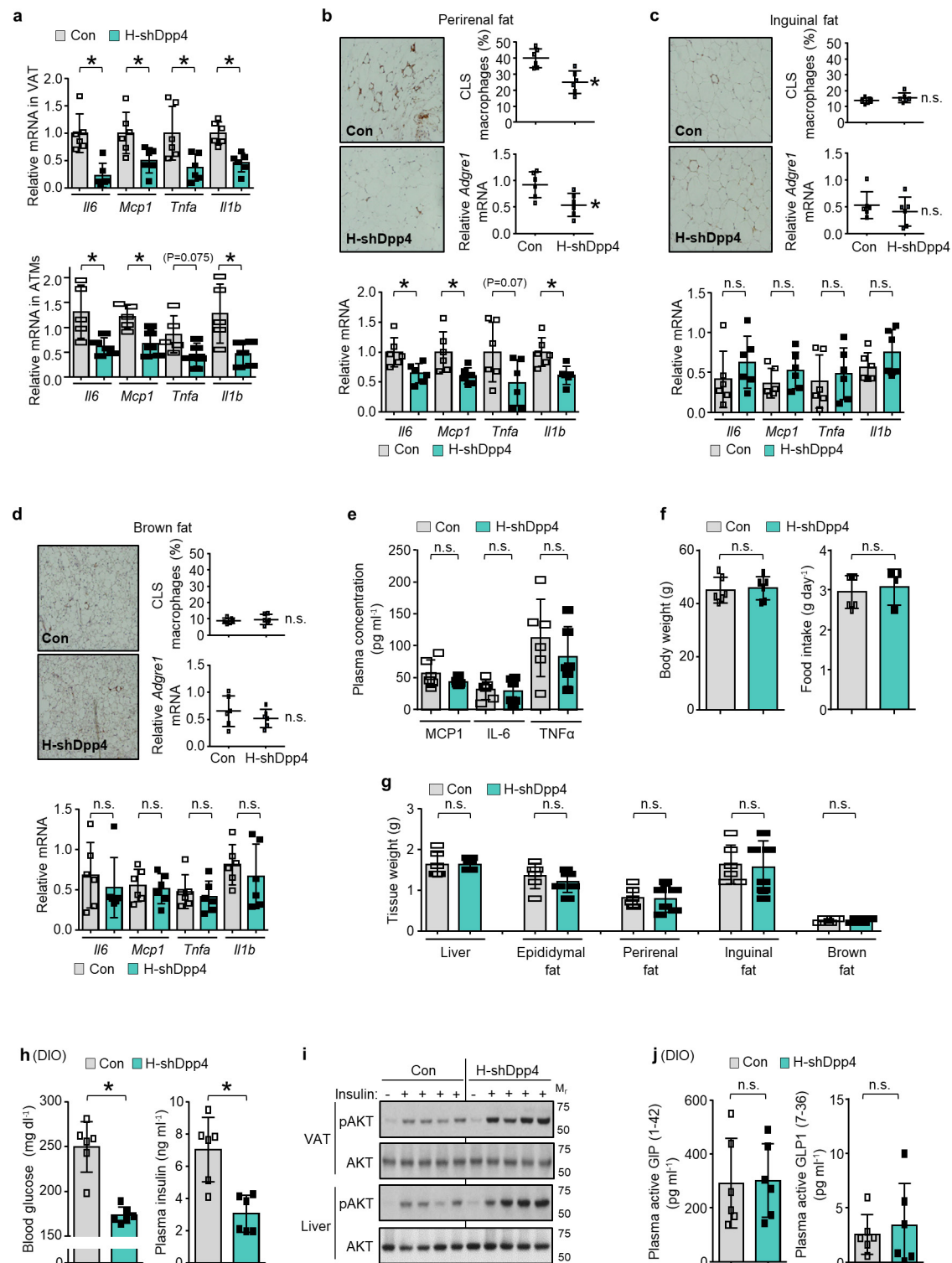
fractionation of DIO mice plasma using gel-filtration FPLC; vertical grey bar depicts peak of activity shown in f. f. Obese mouse SVF cells were incubated with medium containing 10% lean or DIO mouse plasma or the indicated FPLC fractions from e and assayed for *Mcp1* mRNA. n.d., *Mcp1* mRNA not detected. Arrows indicate the fractions that were selected for LC-MS/MS analysis. g, LC-MS/MS normalized spectral counts corresponding to DPP4 in the FPLC fractions from f. h, DPP4 activity in the plasma of lean and DIO mice. *n* = 5 mice per group; mean \pm s.e.m.; **P* < 0.05. i, SVF cells from DIO mice were incubated for 4 h with medium containing 10% (v/v) lean or DIO mouse plasma that was pre-treated for 1 h with or without 10 μ M DPP4 inhibitor KR62436. The cells were then assayed for *Il6* mRNA. *n* = 3 technical replicates per group; mean \pm s.e.m.; **P* < 0.05. j, VAT from the mice in Extended Data Fig. 1f, g was assayed for *Dpp4* mRNA (*n* = 4). Data in a–c and h were analysed by two-tailed Student's *t*-test; data in i and j were analysed by one-way ANOVA (g, h); data in d were analysed by two-way ANOVA.



Extended Data Figure 3 | Restoration of DPP4 in livers of H-CaMKII(KO) mice abrogates suppression of VAT inflammation in DIO mice; ATF4 ChIP of the *Dpp4* gene; and AAV8-H1-shDpp4 treatment lowers hepatic DPP4.

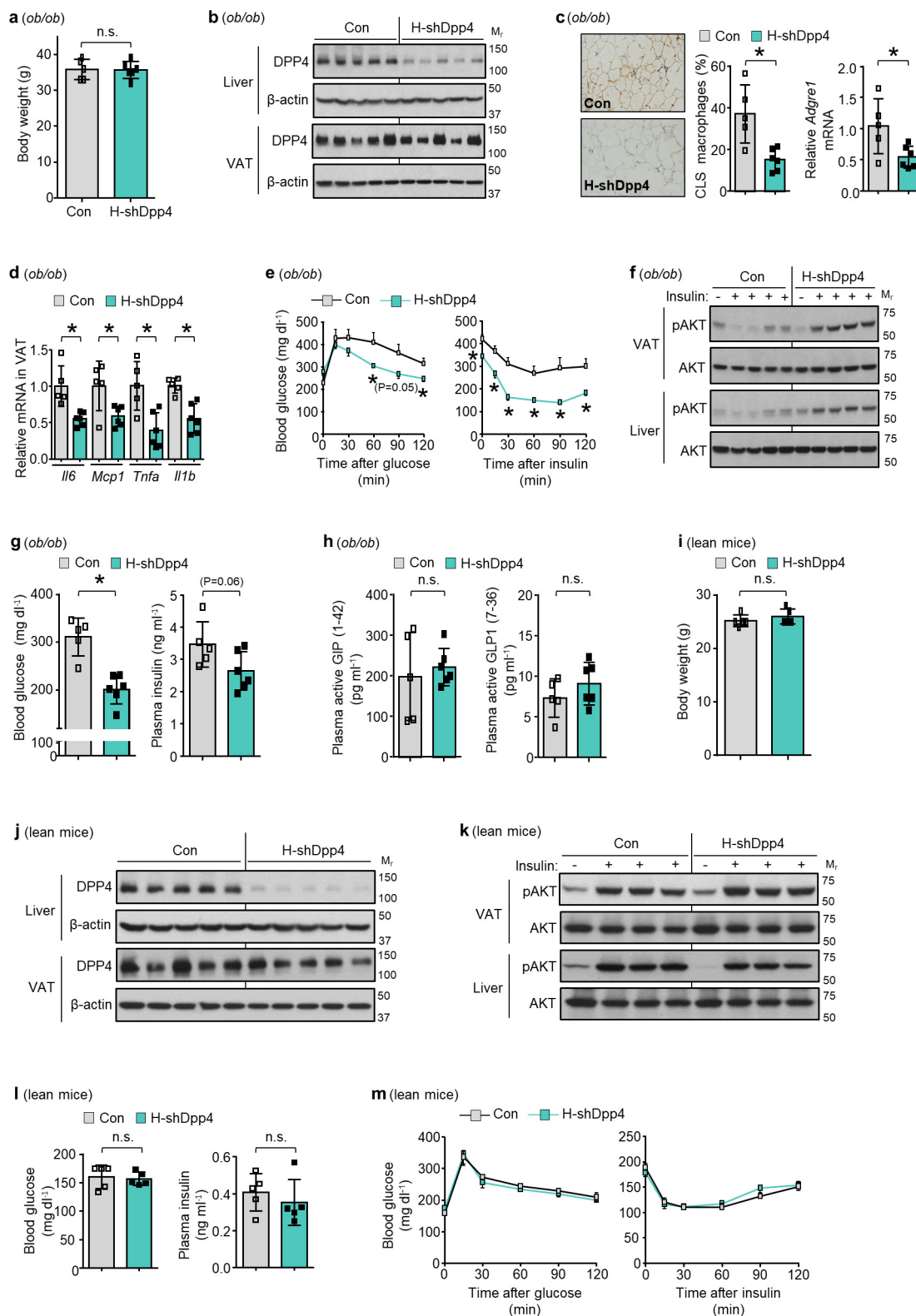
Wild-type and H-CaMKII(KO) mice and a third group in which *Camk2g^{fl/fl}* mice were injected intravenously with adeno-*Dpp4* together with the AAV8-TBG-*cre* (H-CaMKII(KO) + adeno-*Dpp4*) were analysed as follows. **a**, Body weight, plasma DPP4 activity and liver and VAT DPP4 protein. **b**, Representative images of VAT immunostained with F4/80 antibody, with quantification of CLS macrophages and *Tnfa* and *Mcp1* mRNA in VAT. Note that the first two groups of mice received adeno-*lacZ* instead of adeno-DPP4. In **a** and **b**, $n = 3-4$ mice per group; mean \pm s.e.m.; * $P < 0.05$ by one-way ANOVA; n.s., non-significant. For gel source data, see Supplementary Fig. 1. **c**, Top, ChIP was performed with liver extracts from the indicated groups of mice using anti-ATF4 or control IgG antibodies. The region spanning the predicted ATF4-binding sequence in exon 1 of *Dpp4* was amplified by RT-qPCR and normalized to the values obtained from the input.

$n = 3$ ChIP assays for wild-type, H-CaMKII(KO), H-CaMKII(KO) + adeno-*Atf4*; $n = 1$ for H-ATF4(KO); $n = 10$ for control IgG; mean \pm s.e.m.; ** $P < 0.05$ by one-way ANOVA for groups 1–3. Bottom, per cent input for ATF4 ChIP using liver extracts from the indicated DIO mice and PCR primers for a region in the *Dpp4* gene that does not contain a consensus sequence for ATF4 binding. $n = 3$ ChIP assays for wild type, H-CaMKII(KO), H-CaMKII(KO) + adeno-*Atf4*; $n = 1$ for H-ATF4(KO); $n = 10$ for control IgG; mean \pm s.e.m.; n.s., non-significant by one-way ANOVA. **d**, **e**, Sixteen-week-old mice previously fed with the DIO diet for 10 weeks were injected intravenously with hepatocyte-specific AAV8-H1-shDpp4 (H-shDpp4) or control AAV8-H1 vector (con). After four additional weeks on DIO diet, the mice were analysed as follows. **d**, DPP4 immunoblot, with densitometric quantification of DPP4 protein in the indicated tissues; representative of three independent experiments. **e**, *Dpp4* mRNA in liver and VAT. In **d** and **e**, $n = 5$ mice per group; mean \pm s.e.m.; * $P < 0.05$; n.s., non-significant by two-tailed Student's *t*-test. See Supplementary Fig. 1 for gel source data.



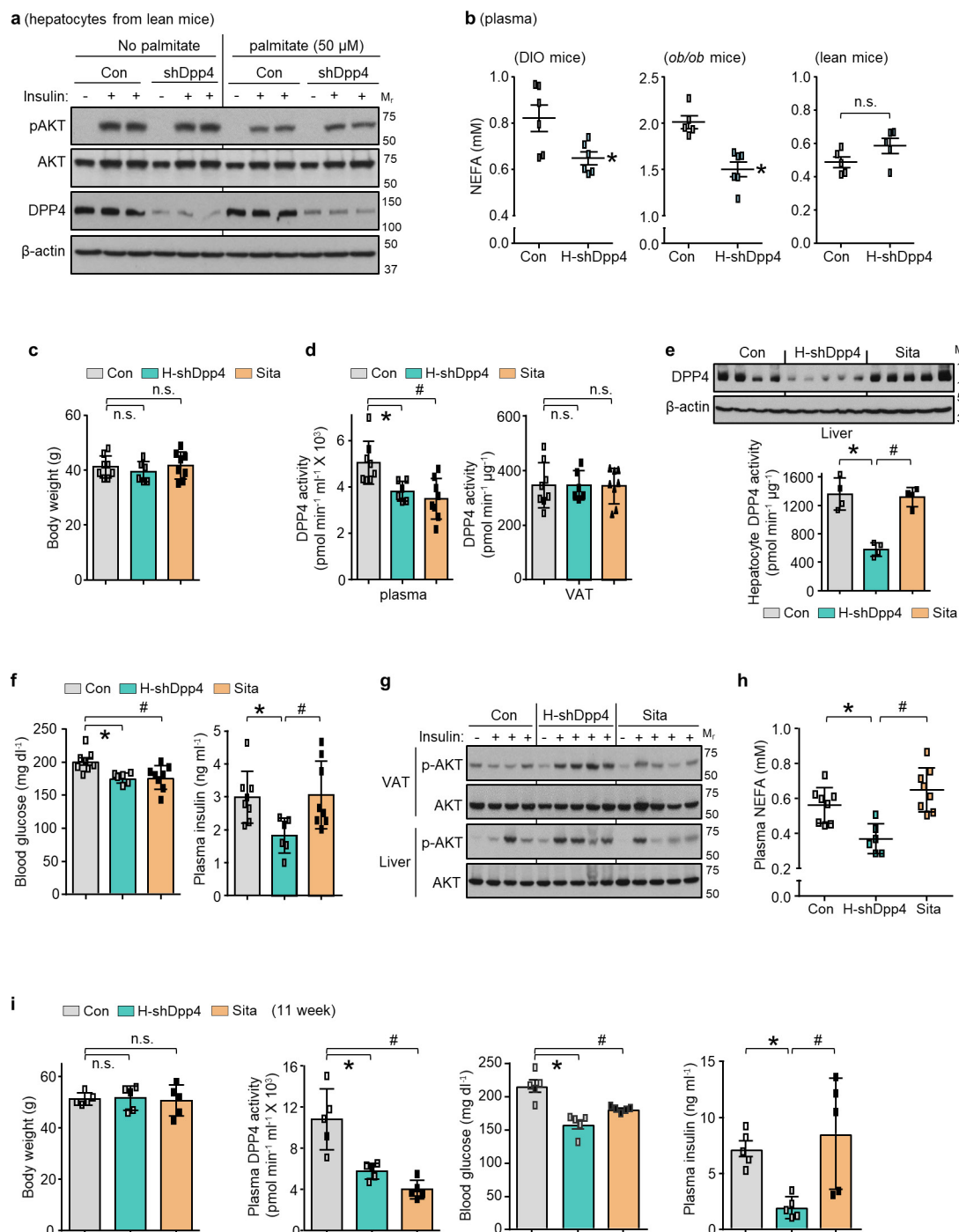
Extended Data Figure 4 | Silencing of DPP4 in liver suppresses VAT inflammation and improves metabolism without increasing plasma incretins. **a**, Top, 16-week-old mice previously fed with the DIO diet for 10 weeks were injected intravenously with AAV8-H1-shDpp4 (H-shDpp4) or control AAV8-H1 vector (con). After four weeks, the mice were assayed for *Il6*, *Mcp1*, *Tnfa* and *Il1b* mRNA in VAT. Bottom, control and H-shDpp4-treated mice similar to those above were analysed nine days after adenovirus injections for *Il6*, *Mcp1*, *Tnfa* and *Il1b* mRNA in ATMs. **b–g**, H-shDpp4 and control mice were analysed after four weeks. **b–d**, Representative images of adipose tissue immunostained with F4/80 antibody, with quantification of CLS macrophages, *Adgre1* and

inflammatory gene mRNA expression in perirenal, inguinal and brown fat. **e**, Plasma MCP1, IL6 and TNF α . **f**, Body weight and food intake. **g**, Weights of liver and the indicated adipose tissue depots. In **a–g**, $n = 5–6$ mice per group; mean \pm s.e.m.; * $P < 0.05$; n.s., non-significant by two-tailed Student's t -test. **h–j**, Mice similar to those in **b–g** were assayed. **h**, Blood glucose and plasma insulin. **i**, p-AKT and total AKT in VAT and liver after insulin injection into the portal vein. $n = 1$ PBS-injected, $n = 4$ insulin-injected mice per group; blots are representative of three independent experiments. Gel source data are shown in Supplementary Fig. 1. **j**, Plasma active GIP (1–42) and GLP1 (7–36). In **h** and **j**, $n = 6$ mice per group; mean \pm s.e.m.; * $P < 0.05$; n.s., non-significant by two-tailed



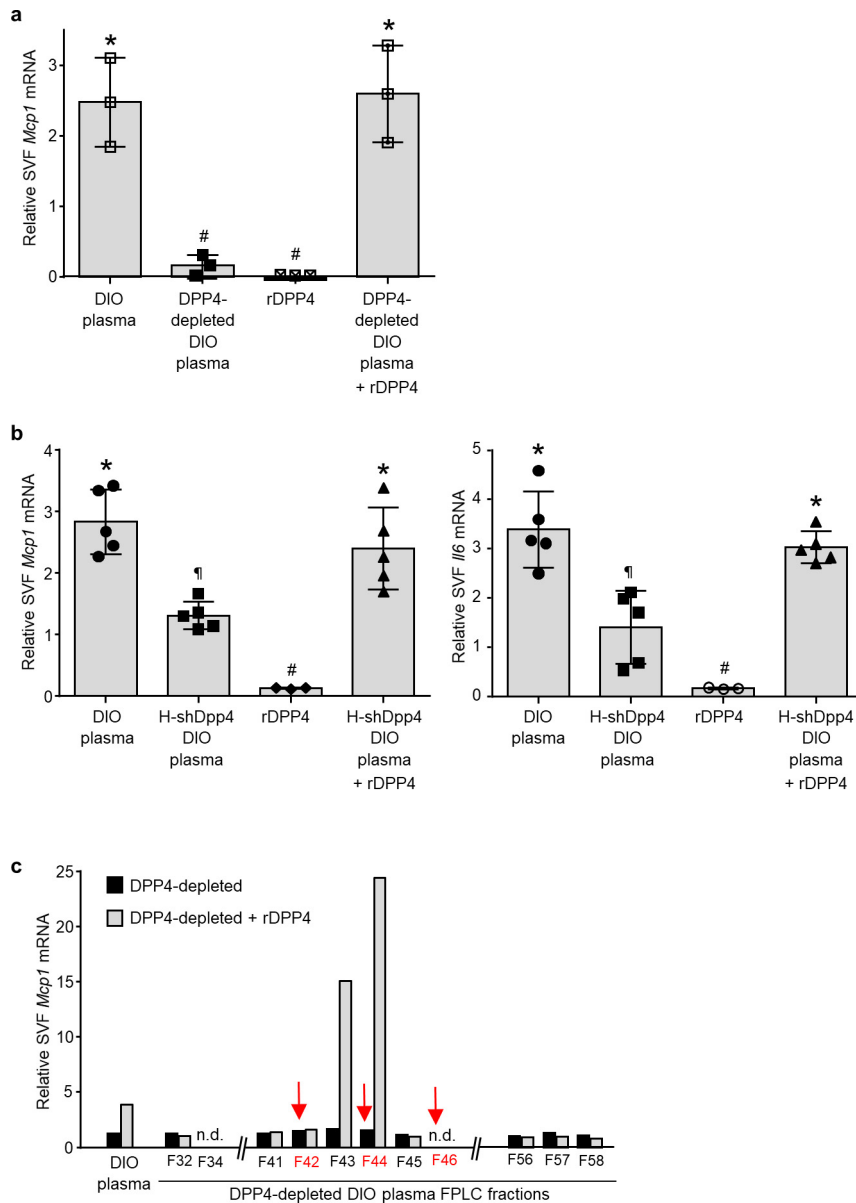
Extended Data Figure 5 | Hepatocyte-specific silencing of DPP4 improves glucose metabolism in *ob/ob* mice without increasing plasma incretins and does not affect VAT inflammation or glucose metabolism in lean mice. **a–h**, Five-week-old chow-fed *ob/ob* mice were injected intravenously with AAV8-H1-shDpp4 (H-shDpp4) or AAV8-H1-control (con), and were assayed four weeks later. **a**, Body weight. **b**, Immunoblot of liver and VAT DPP4. **c**, Representative images of VAT immunostained with F4/80 antibody, with quantification of CLS macrophages and *Adgre1* mRNA in VAT. **d**, *Il6*, *Mcp1*, *Tnfa* and *Il1b* mRNA in VAT. **e**, Blood glucose after challenge with intraperitoneal glucose or insulin. **f**, p-AKT and total AKT in VAT and liver extracts after portal vein insulin injection. **g**, Blood glucose and plasma

insulin 5 h after food withdrawal. **h**, Plasma active GIP (1–42) and GLP1 (7–36). **i–m**, Sixteen-week-old chow-fed wild-type lean mice were injected intravenously with AAV8-H1-shDpp4 (H-shDpp4) or AAV8-H1-control (Con) and were analysed as follows. **i**, Body weight. **j**, Immunoblot of DPP4 in liver and VAT. **k**, p-AKT and total AKT in VAT and liver extracts after portal vein insulin injection. **l**, Blood glucose and plasma insulin 5 h after food withdrawal. **m**, Blood glucose after challenge with intraperitoneal glucose or insulin. In all panels, $n = 5–6$ mice per group; mean \pm s.e.m.; n.s., non-significant by two-tailed Student's *t*-test. For gel source data, see Supplementary Fig. 1.



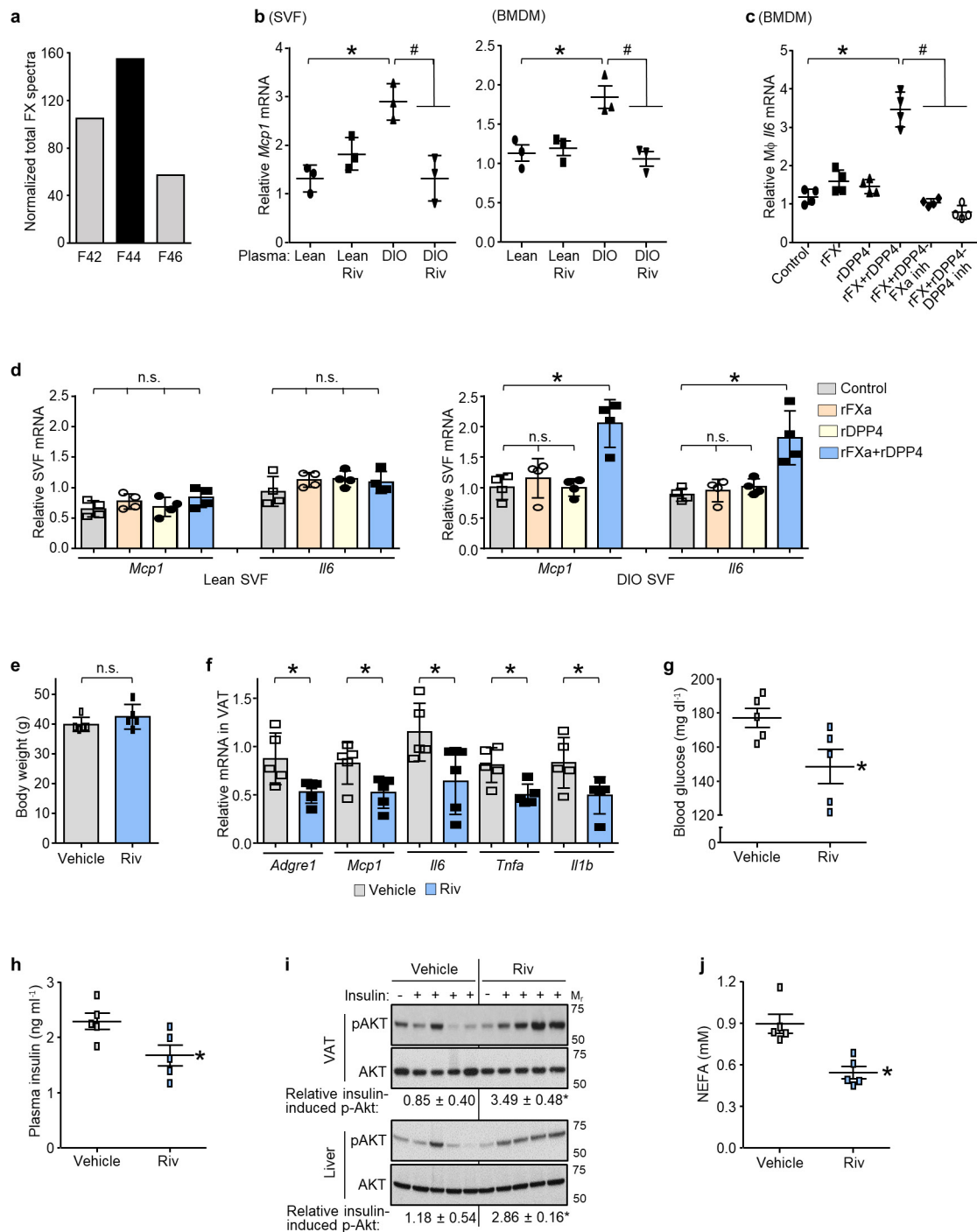
Extended Data Figure 6 | Effect of DPP4 silencing on insulin-induced p-AKT in primary hepatocytes, on NEFA in obese or lean mice and on metabolism in comparison with sitagliptin in obese mice. **a**, Wild-type or DPP4-silenced primary hepatocytes were treated with or without 50 μ M palmitate for 10 h with the last 5 h in serum-free medium, and then stimulated with 100 nM of insulin for 5 min. p-AKT and total AKT and DPP4 were assayed by immunoblot. The data are representative of two independent experiments. **b**, Plasma samples from the following mice were assayed for non-esterified fatty acids (NEFA) four weeks after intravenous injection with AAV8-H1-shDpp4 (H-shDpp4) or control AAV8-H1 vector (con): 16-week-old mice fed the DIO diet for the last 10 weeks, 5-week-old chow-fed *ob/ob* mice, and 6-week-old chow-fed wild-type lean mice. $n = 5$ –6 mice per group; mean \pm s.e.m.; * $P < 0.05$ and n.s., non-significant by two-tailed Student's *t*-test. **c**–**h**, After 10 weeks on high-fat diet,

control DIO mice, H-shDpp4 DIO mice and DIO mice were administered sitagliptin (sita) in drinking water to achieve a dose of ~ 30 –45 mg/kg/day. After 4 weeks of treatment, the mice were analysed as follows. **c**, Body weight. **d**, Plasma and VAT DPP4 activity. **e**, Hepatic DPP4 immunoblot and hepatocyte DPP4 activity. **f**, Blood glucose and plasma insulin 5 h after food withdrawal. **g**, p-AKT and total AKT in VAT and liver after portal vein insulin injection. **h**, Plasma non-esterified fatty acids (NEFA). For all quantified data panels except **e**, $n = 6$ –8 mice per group; for **e**, $n = 4$ mice per group; mean \pm s.e.m.; * $P < 0.05$ and n.s., non-significant by one-way ANOVA. For gel source data, see Supplementary Fig. 1. **i**, DIO mice were treated with AAV8-H1-con, AAV8-H1-shDpp4 or sitagliptin (+ AAV8-H1-con) exactly as above, except the mice were analysed after 11 weeks of treatment instead of after 4 weeks. $n = 5$ mice per group; mean \pm s.e.m.; * $P < 0.05$ and n.s., non-significant by one-way ANOVA.



Extended Data Figure 7 | DPP4 requires a plasma factor to induce inflammation in SVF. **a**, *Mcp1* mRNA was assayed in SVF cells that were incubated for 4 h with medium containing 10% (v/v) DIO mouse plasma; DIO mouse plasma immunodepleted of DPP4; recombinant DPP4 (rDPP4) alone; or rDPP4 plus DIO mouse plasma immunodepleted of DPP4. $n = 3$ technical replicates per group; mean \pm s.e.m. **b**, As **a**, except plasma from H-shDPP4 DIO mice were used instead of DPP4-depleted obese mouse plasma, and both *Mcp1* and *Il6* were assayed ($n = 5$

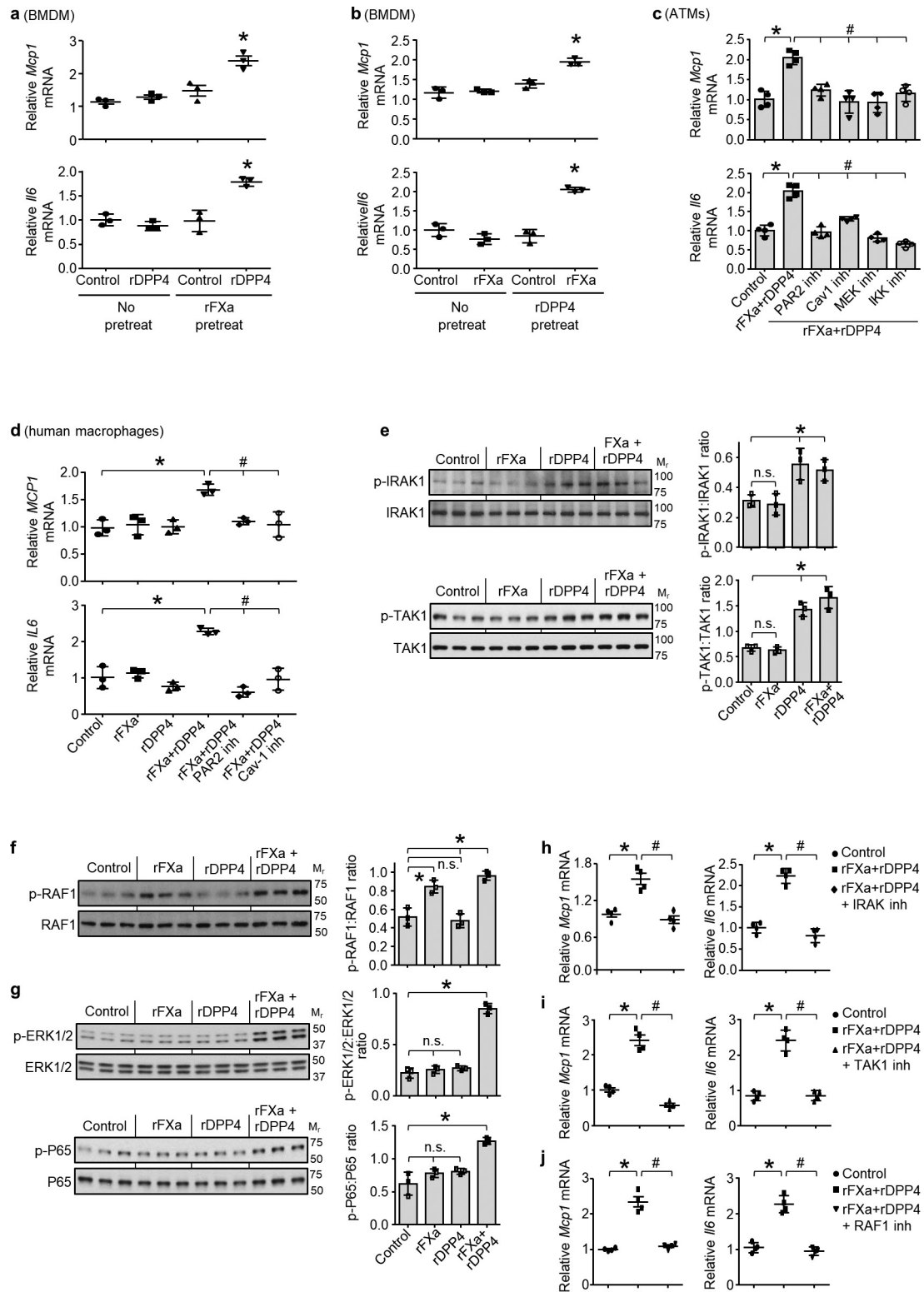
technical replicates per group; mean \pm s.e.m.). In **a** and **b**, groups with different symbols are different from each other ($P < 0.05$ by one-way ANOVA). **c**, DPP4-depleted plasma from DIO mice was fractionated by Superdex-200 FPLC. Each fraction, as well as unfractionated DIO mouse plasma, was incubated with SVF cells with or without rDPP4, followed by assay of *Mcp1* mRNA. n.d., *Mcp1* mRNA not detected. The fraction numbers in red (arrows) were selected for LC-MS/MS analysis. The data are from a single experiment.



Extended Data Figure 8 | DPP4 and FXa synergistically promote VAT inflammation, and inhibition of FXa improves glucose metabolism.

a, LC-MS/MS normalized spectral counts corresponding to FX in the indicated FPLC fractions from Extended Fig. 7c. **b**, SVF cells or BMDMs were pre-incubated for 1 h with or without 10 μ M FXa inhibitor rivaroxaban (riv) and then incubated with medium containing 10% (v/v) plasma from lean or DIO mice and assayed for *Mcp1* mRNA. $n = 3$ technical replicates per group; mean \pm s.e.m.; $^{**}P < 0.05$ by two-way ANOVA. **c**, BMDMs were pretreated with or without 10 μ M FXa inhibitor rivaroxaban (riv) or 10 μ M DPP4 inhibitor KR62436, followed by incubation for 4 h with rFX or rDPP4 alone or both together. *Il6* mRNA was then quantified. $n = 4$ technical replicates per group; mean \pm s.e.m.; $^{*}P < 0.05$ by one-way ANOVA. **d**, SVF cells from lean or DIO mice were treated with rFX or rDPP4 alone or in combination for 4 h, and then *Mcp1*

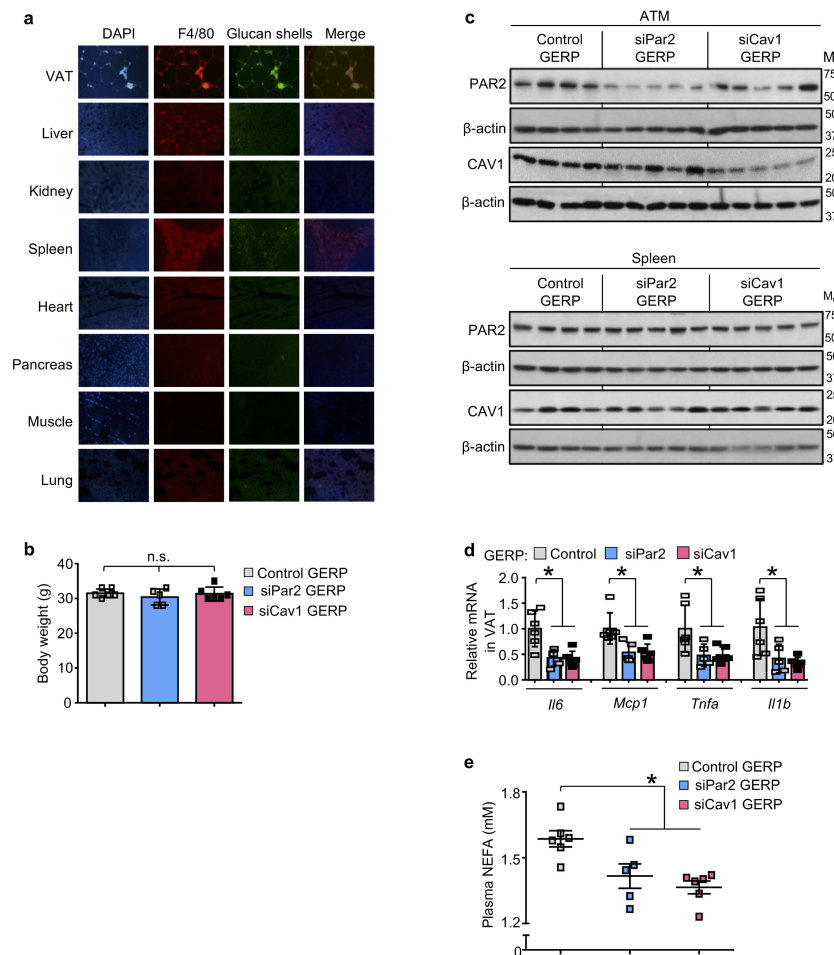
and *Il6* mRNA levels were assayed ($n = 4$ technical replicates per group; mean \pm s.e.m.; $^{*}P < 0.05$ and n.s., non-significant by one-way ANOVA). **e–j**, The control and rivaroxaban-treated mice from Fig. 3b, c were assayed. **e**, Body weight. **f**, *Adgre1*, *Mcp1*, *Il6*, *Tnfa* and *Il1b* mRNA in VAT. **g**, **h**, Blood glucose and plasma insulin 5 h after food withdrawal. **i**, p-AKT and total AKT in VAT and liver after portal vein insulin injection. **j**, Plasma non-esterified fatty acids (NEFA). In **e–h** and **j**, $n = 5$ mice per group; mean \pm s.e.m.; $^{*}P < 0.05$ and n.s., non-significant by two-tailed Student's *t*-test. In **i**, the mean fold increases of the plus-insulin values relative to the minus-insulin value, based on the densitometric ratio of p-AKT:total AKT, are shown below the blots ($n = 1$ PBS-injected and $n = 4$ insulin injected mice per group; mean \pm s.e.m.; $^{*}P < 0.05$). For gel source data, see Supplementary Fig. 1.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | DPP4 and FXa synergistically induce inflammatory signalling pathways in macrophages. **a**, BMDMs were pre-treated for 4 h with or without 0.25 U rFXa, medium was then removed, cells were washed and treated with 10 μ M rivaroxaban (riv) to inhibit residual FXa activity. The cells were then incubated for 4 h with or without 70 ng rDPP4 and assayed for *Mcp1* and *Il6* mRNA. **b**, BMDMs were pre-treated for 4 h with or without 70 ng rDPP4, medium was then removed, cells were washed and treated with 10 μ M KR62436 to inhibit residual DPP4 activity. The cells were then incubated for 4 h with or without 0.25 U rFXa and assayed for *Mcp1* and *Il6* mRNA. In **a** and **b**, $n = 3$ technical replicates per group; mean \pm s.e.m.; * $P < 0.05$ versus all other groups by one-way ANOVA. **c**, ATMs were pre-treated for 1 h with or without 10 μ M of the PAR2 inhibitor GB83, 25 μ M of the CAV1 inhibitor daidzein, 10 μ M of the MEK inhibitor PD98059 or 10 μ M of the IKK inhibitor PS-1145. The cells were then incubated for 4 h with or without rFXa and rDPP4 and assayed for *Mcp1* and *Il6* mRNA. $n = 4$ technical replicates per group; mean \pm s.e.m.; * $P < 0.05$ by one-way ANOVA. **d**, Human monocyte-derived macrophages were pre-treated for 1 h with or without 10 μ M PAR2

inhibitor (GB83) or 25 μ M CAV1 inhibitor (daidzein) and then incubated for 4 h with or without rFXa, rDPP4 or both and assayed for *MCP1* and *IL6* mRNA. $n = 3$ technical replicates; mean \pm s.e.m.; * $P < 0.05$ versus all other groups by one-way ANOVA. **e–g**, Bone marrow-derived macrophages were incubated for 10 min with or without 0.25 U rFXa, 70 ng rDPP4 or rFXa plus rDPP4 and then assayed by immunoblot for phosphorylated and total IRAK1, TAK1, RAF1, ERK1/2 and P65, followed by densitometric quantification. The data are representative of two (**e**, **f**) or three (**g**) independent experiments. For all panels, $n = 3$ technical replicates per group; mean \pm s.e.m.; * $P < 0.05$ versus other groups and n.s., not significant by one-way ANOVA. **h–j**, Bone marrow-derived macrophages were pretreated for 1 h with or without the following inhibitors. **h**, IRAK-1/4 inhibitor-I, 0.5 μ M. **i**, TAK1 MAPKKK inhibitor Z-7-oxozeaenol, 0.1 μ M. **j**, RAF1 inhibitor GW5074, 1 μ M. Cells were then incubated for 4 h with or without rFXa plus rDPP4 and assayed for *Mcp1* and *Il6* mRNA. $n = 4$ technical replicates per group; mean \pm s.e.m.; * $P < 0.05$ by one-way ANOVA.



Extended Data Figure 10 | Supporting data for GERP experiments.

a, Five-week-old chow-fed *ob/ob* mice were treated daily by intraperitoneal injection of FITC-labelled glucan shells for 12 days and then assayed by DAPI staining, F4/80 immunofluorescence and glucan shell fluorescence in the indicated tissues. **b–e**, Five-week-old chow-fed *ob/ob* mice were injected intraperitoneally once daily for 12 days with 0.2 mg GERPs containing scrambled siRNA (control), *Par2* siRNA or *Cav1* siRNA.

The mice were analysed 24 h after the last injection. **b**, Body weight. **c**, Immunoblots of PAR2 and CAV1 in ATMs and splenic extracts. **d**, *Il6*, *Mcp1*, *Tnfa* and *Il1b* mRNA in VAT. **e**, Plasma non-esterified fatty acids (NEFA). In **b**, **d** and **e**, $n = 5–6$ mice per group; * $P < 0.05$ and n.s., non-significant by two-tailed Student's *t*-test for each siRNA GERP versus control GERP. For gel source data, see Supplementary Fig. 1.

The protein histidine phosphatase LHPP is a tumour suppressor

Sravanth K. Hindupur¹, Marco Colombi¹, Stephen R. Fuhs², Matthias S. Matter³, Yakir Guri¹, Kevin Adam², Marion Cornu¹, Salvatore Piscuoglio³, Charlotte K. Y. Ng³, Charles Betz¹, Dritan Liko¹, Luca Quagliata³, Suzette Moes¹, Paul Jenoe¹, Luigi M. Terracciano³, Markus H. Heim⁴, Tony Hunter² & Michael N. Hall¹

Histidine phosphorylation, the so-called hidden phosphoproteome, is a poorly characterized post-translational modification of proteins^{1,2}. Here we describe a role of histidine phosphorylation in tumorigenesis. Proteomic analysis of 12 tumours from an mTOR-driven hepatocellular carcinoma mouse model revealed that NME1 and NME2, the only known mammalian histidine kinases, were upregulated. Conversely, expression of the putative histidine phosphatase LHPP was downregulated specifically in the tumours. We demonstrate that LHPP is indeed a protein histidine phosphatase. Consistent with these observations, global histidine phosphorylation was significantly upregulated in the liver tumours. Sustained, hepatic expression of LHPP in the hepatocellular carcinoma mouse model reduced tumour burden and prevented the loss of liver function. Finally, in patients with hepatocellular carcinoma, low expression of LHPP correlated with increased tumour severity and reduced overall survival. Thus, LHPP is a protein histidine phosphatase and tumour suppressor, suggesting that deregulated histidine phosphorylation is oncogenic.

Liver cancer is the second leading cause of cancer-related deaths globally. Hepatocellular carcinoma (HCC) represents approximately 90% of primary liver cancer cases³. Recent studies indicate that almost 50% of HCC cases display aberrant PI3K–AKT–mTOR signalling⁴, including loss of the tumour suppressors PTEN, TSC1, and TSC2³ (Fig. 1a). We generated an HCC mouse model by liver-specific deletion of PTEN and TSC1, using the neonatally expressed albumin promoter as a *cre* driver (see Methods). We hereafter refer to this model as liver-specific double-knockout (L-dKO) mice. L-dKO mice invariably exhibited hepatomegaly and advanced liver tumours at 6 and 20 weeks of age, respectively (Fig. 1b and Extended Data Fig. 1a). Histopathological analysis (Fig. 1b) and molecular classification based on mRNA expression of the liver cancer markers *CD133* (also known as *Prom1*), *CD90* (*Thy1*), *Cd44* and *Aldh1a1*) revealed that the L-dKO liver tumours indeed mimicked poorly differentiated human HCC (Extended Data Fig. 1b), consistent with other studies^{5,6}.

We performed quantitative proteomic analysis on three tumours each from four 20-week-old, L-dKO mice (12 tumours in total; see Methods). Liver protein extracts from six *cre*-negative age- and sex-matched littermates were pooled in equal amounts and used as a control (Extended Data Fig. 1c). We identified approximately 4,500 proteins per tumour (Extended Data Fig. 1d). Of the 3,147 proteins that were detected in a minimum of 10 tumours, 433 proteins were upregulated and 262 proteins were downregulated compared to the control (Fig. 1c). We manually curated a list of known or putative phosphatases and kinases significantly deregulated in a minimum of ten tumours (Extended Data Fig. 1e). Particularly intriguing in this list were the upregulated histidine kinases NME1 and NME2 (also known as NDPKA and NDPKB, respectively) (Fig. 1d), and the downregulated

putative histidine phosphatase LHPP (phospholysine phosphohistidine inorganic pyrophosphate phosphatase).

LHPP has been purified from bovine liver and shown to hydrolyse P–N bonds in synthetic substrates *in vitro*^{7–9}, but is otherwise poorly characterized. LHPP is evolutionarily conserved from worm to human (overall 28% identity) (Extended Data Fig. 2a), but only weakly homologous to the two defined mammalian histidine phosphatases PHPT1 (8.1% identity) and PGAM5 (6.4% identity)^{10–12} (Extended Data Fig. 2b). We note that PHPT1 and PGAM5 levels were unchanged in tumours from L-dKO mice (Fig. 1d). NME1 and NME2 are homologous proteins (88% identical) that function as histidine kinases as either homo- or hetero-oligomers. NME1 and NME2 are the only mammalian protein histidine kinases reported so far^{13–15}. Immunoblot analysis confirmed that LHPP was downregulated and NME1 and NME2 were upregulated (Fig. 2a and Extended Data Fig. 3a) in L-dKO tumours, as initially observed by proteomic analysis. Immunoblot analysis of S6 protein phosphorylated at Ser240 and Ser244 (S6-pS240/244) and AKT phosphorylated at Ser473 (AKT-pS473) confirmed high mTORC1 and mTORC2 activity, respectively, in L-dKO liver tissue and tumours (Fig. 2a, b). To determine whether the loss of LHPP expression coincided with tumorigenesis, we examined whole liver tissue (6, 12 and 16 weeks) and isolated tumours (20 weeks) from L-dKO mice at different ages and hence at different stages of tumour development (Fig. 2b). Age-matched *cre*-negative littermates served as controls. LHPP expression was mildly reduced in the livers of L-dKO mice at 16 weeks, at which age macroscopic tumours were first visible. As observed previously, LHPP was strongly decreased in tumours of 20-week-old mice. As expected, mTOR signalling was high at all ages. Thus, the reduction in LHPP expression coincided with tumour development, but did not correlate with mTOR activity. Hepatic NME1 and NME2 expression decreased in control mice in an age-dependent manner, but remained high in L-dKO liver tissue and tumours. Thus, increased NME1 and NME2 expression in L-dKO mice correlated with increased mTOR activity rather than with tumour development (Fig. 2b), suggesting that mTOR may control NME1 and NME2 expression. To determine whether loss of LHPP expression was specific to tumours, we performed immunoblot analysis on dissected tumour and non-tumour liver tissue from 20-week L-dKO mice. LHPP expression was reduced (or absent) in the tumour, compared to the non-tumour liver tissue (Extended Data Fig. 3b). Immunohistochemistry on liver from 20-week L-dKO mice confirmed that LHPP was strongly downregulated (undetectable) in tumours compared to non-tumour tissue (Extended Data Fig. 3c). Like mTOR activity, NME1 and NME2 expression was similarly upregulated in both tumour and non-tumour L-dKO liver tissue compared to control liver tissue (Extended Data Fig. 3b, c), again indicating that the expression of NME1 and NME2 correlated with mTOR activity rather than tumorigenicity. These findings suggest that LHPP

¹Biozentrum, University of Basel, 4056 Basel, Switzerland. ²Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA. ³Institute of Pathology, University Hospital Basel, 4031 Basel, Switzerland. ⁴Department of Biomedicine, University Hospital Basel, 4031 Basel, Switzerland.

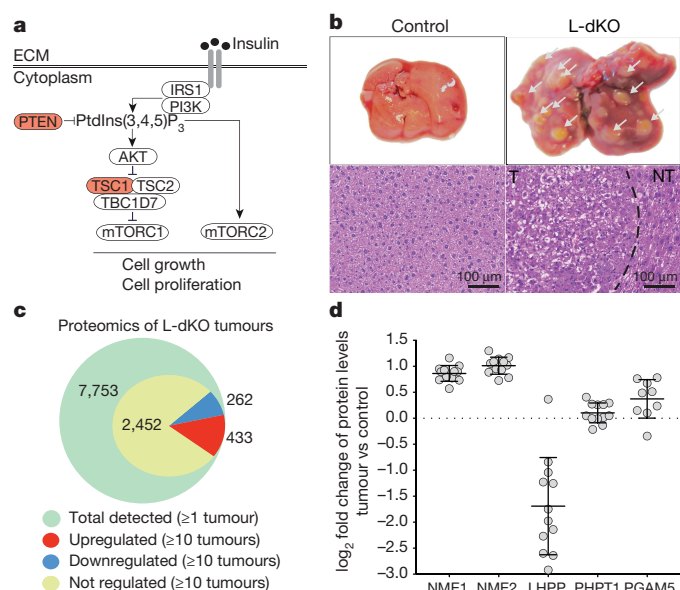


Figure 1 | Proteomics on L-dKO tumours. **a**, Diagram of mTOR signalling showing TSC and PTEN tumour suppressors. **b**, Top, representative images of whole livers from 20-week-old L-dKO (tumours indicated with arrowheads) and control mice. Bottom, representative photomicrographs of haematoxylin and eosin (H&E)-stained liver from control mice, tumour (T) and adjacent non-tumour (NT) tissue from L-dKO mice (20 weeks). Similar results were obtained in tumours from nine mice. **c**, Number of tumour proteins detected and quantified by mass spectrometry. A total of 7,753 proteins were quantified in all tumours ($n = 12$). 3,147 proteins were quantified in a minimum of 10 tumours, of which 433 proteins were upregulated (red), 262 were downregulated (blue), and 2,452 were unchanged (yellow). ANOVA-based two-sample t -test with a false discovery rate (FDR) of 2% was used to determine regulation. **d**, Mass spectrometry-determined fold change in independent L-dKO tumours ($n = 12$ for NME1, NME2, LHPP and PHPT1; $n = 9$ for PGAM5) compared to control livers ($n = 6$); data are mean \pm s.d. PHPT1 and PGAM5 are two defined mammalian histidine phosphatases.

inactivation, in a background of constitutively high NME1 and NME2 activity in L-dKO liver, is a key event in tumorigenesis.

The above findings suggest that the expression of NME1 and NME2, but not LHPP, is controlled by mTOR. To investigate this further, we acutely treated (24 h) L-dKO mice with the mTORC1 inhibitor rapamycin or the pan-mTOR inhibitor INK128. INK128 treatment reduced NME1 and NME2 expression in L-dKO tumours, but had no effect on LHPP (Extended Data Fig. 3d). Rapamycin had no effect on the level of NME1, NME2 or LHPP (data not shown). Thus, the expression of NME1 and NME2 seems to be mTORC2 dependent, whereas LHPP expression is mTOR independent.

The paired upregulation of NME1 and NME2 (histidine kinases) and downregulation of LHPP (putative histidine phosphatase) prompted us to investigate protein histidine phosphorylation in L-dKO tumours^{1,16}. Phosphohistidine (pHis) exists in two isomers, 1-pHis and 3-pHis, depending on the position of the phospho-acceptor nitrogen in the histidine imidazole ring (N1 and N3 positions, respectively)^{2,17}. Unlike the phosphoester (P–O) bond in other phosphoamino acids, the phosphoramidate (P–N) bond in phosphohistidine is heat- and acid-labile, thus making pHis difficult to detect in biological samples¹⁸. This difficulty has been largely circumvented by the recent development of monoclonal antibodies that specifically recognize 1-pHis or 3-pHis². The new antibodies were used to assess histidine phosphorylation (1-pHis and 3-pHis) in L-dKO tumours and control liver (Fig. 3a) (see Methods). The immunoblot signals of both 1-pHis and 3-pHis were high in tumours compared to control tissue (Fig. 3a, b). Heating sample-buffer-solubilized lysates at 95 °C for 10 min before SDS–PAGE caused

dephosphorylation, indicating that the signals detected in the unheated samples are indeed histidine-phosphorylated proteins (Fig. 3a). As reported previously with cultured cells², multiple 3-pHis-positive proteins but only one or a few 1-pHis proteins were detected in tumours and liver tissue. This could be due to the lower thermodynamic stability of 1-pHis than of 3-pHis^{19,20}, or the existence of fewer proteins phosphorylated at the N1 position. The major band observed with the 1-pHis antibody is likely to be co-migrating NME1-pHis118 and NME2-pHis118 (Fig. 3a), because it displays an appropriate molecular mass (18 kDa) for NME1 and NME2 and was lost upon NME1 and NME2 knockdown (data not shown). NME1 and NME2 normally autophosphorylate position N1 of His118, and then transfer this phosphate to histidine (N1 or N3) in target proteins (Extended Data Fig. 4a). Thus, consistent with increased levels of NME1-pHis118 and NME2-pHis118 and decreased LHPP levels, histidine phosphorylation was significantly increased in L-dKO tumours.

To identify proteins in which histidine phosphorylation is increased in tumours, we used a combination of anti-1-pHis and anti-3-pHis monoclonal antibodies to immunoprecipitate pHis proteins from four L-dKO tumours and control tissue. Subsequent mass spectrometry identified 1,173 and 823 proteins from tumour and control liver tissue lysates, respectively (Supplementary Tables 1 and 2). In total, 260 proteins were more abundant (at least twofold) in tumours compared to control tissue (Supplementary Table 3). Only 18 proteins were more abundant in control tissue (see below for more discussion).

We next investigated whether LHPP is indeed a protein phosphohistidine phosphatase. First, we assayed LHPP phosphohistidine phosphatase activity *in vitro* (see Methods). Recombinant LHPP treatment reduced 3-pHis levels in an electrophoretically fractionated CB1 cell lysate transferred to a membrane (Extended Data Fig. 4b). CB1 is a cell line derived from L-dKO tumours (see Methods) and thus displays high levels of histidine phosphorylation. Second, we assessed LHPP phosphatase activity *in vivo*. Restoration of *Lhpp* expression in CB1 cells, using an adenovirus delivery system, reduced 3-pHis 3.1-fold with negligible effect on 1-pHis, pSer/Thr or pTyr (Fig. 3c, d, Extended Data Fig. 4c). Similar results were obtained with regard to pHis (3.9-fold drop in 3-pHis, no effect on 1-pHis) after introducing mouse *Lhpp* into human SNU449 cells, a liver cancer cell line with low intrinsic LHPP levels (Extended Data Fig. 4d, e). Immunoblot analysis confirmed the expression of recombinant *Lhpp* in both cell lines (Fig. 3c and Extended Data Fig. 4d). Collectively, these results indicate that LHPP is a histidine phosphatase broadly acting on N3-phosphorylated proteins. We note that PGAM5 was previously demonstrated¹² to be a 1-pHis phosphatase acting specifically on NME2-pHis118.

To identify potential LHPP targets, we immunoprecipitated N3-phosphorylated proteins from CB1 cells lacking and re-expressing LHPP. Mass spectrometry revealed nine proteins that were more abundant (at least 1.7-fold) in CB1 cells that lack LHPP (Supplementary Table 4). These included ACLY (ATP citrate lyase), known to be N3-phosphorylated at His760 (see below).

We next examined further the role of the phosphatase LHPP in proliferation and tumorigenesis. Increased expression of LHPP in CB1 and SNU449 cells significantly impaired cell proliferation and, in CB1 cells, reduced hepatosphere formation (Extended Data Fig. 4f, g). Conversely, knocking down the already low level of LHPP expression in SNU449 cells enhanced cell proliferation (Extended Data Fig. 5a). Next, we investigated the effect of LHPP expression on tumour formation in L-dKO mice. Hepatocyte-specific, adeno-associated virus (AAV) expressing either red fluorescent protein (RFP) (AAV/DJ-ALBp-RFP, hereafter referred to as AAV-control) or RFP and murine LHPP (AAV/DJ-ALBp-RFP-2A-mLhpp, hereafter referred to as AAV-LHPP) was injected into the tail vein of 8-week-old L-dKO and littermate control mice (Fig. 3e). Tumour burden was assessed at 20 weeks of age. Hepatic overexpression of LHPP in AAV-LHPP-infected L-dKO mice (20 weeks) was confirmed by immunoblotting (Extended Data Fig. 5b). Macroscopic and histological examination of livers revealed few and

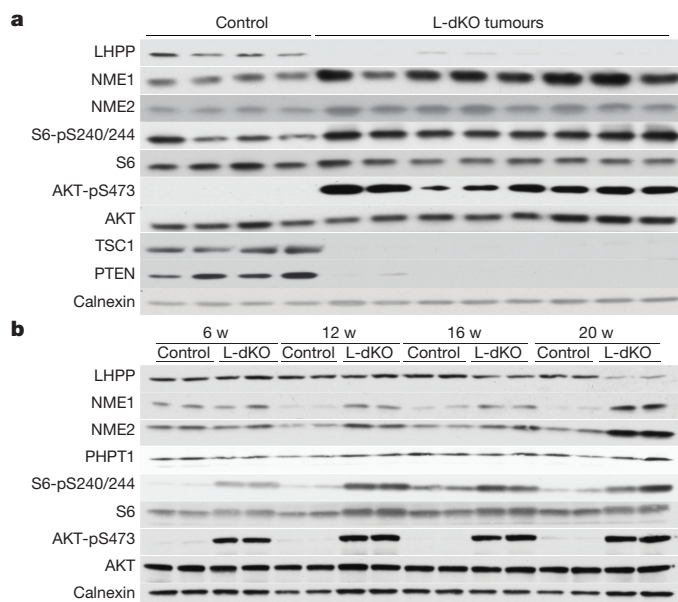


Figure 2 | Loss of LHPP expression is confined to L-dKO tumours.

a, Immunoblot analysis indicates reduced LHPP and increased NME1 and NME2 expression in tumours compared to age-matched littermate control (liver samples from control mice ($n = 4$) and two tumour samples each from four 20-week-old L-dKO mice ($n = 8$)). For full scans, see Supplementary Fig. 1. **b**, Immunoblot analysis indicates reduced LHPP expression (16 and 20 weeks (w)) and increased NME1 and NME2 expression (12, 16 and 20 weeks) in L-dKO liver (6, 12 and 16 weeks) and tumours (20 weeks) compared to age-matched control mice (6, 12 and 16 weeks, $n = 4$, liver tissues from two mice are pooled per lane). For full scans, see Supplementary Fig. 2.

small (micronodular) tumours in AAV-LHPP-infected L-dKO mice (Fig. 3e–g). The few ‘escaper’ tumours that arose in AAV-LHPP-infected L-dKO mice displayed reduced LHPP expression, again suggesting that LHPP downregulation is essential for tumorigenesis. AAV-control-infected L-dKO livers exhibited severe HCC with multiple large neoplastic (macronodular) tumours. Serum levels of alanine amino transferase (ALT), aspartate amino transferase (AST) and lactate dehydrogenase (LDH), well characterized indicators of liver damage, were also reduced in AAV-LHPP-infected L-dKO mice compared to AAV-control-infected mice (Extended Data Fig. 5c). No detrimental effect was observed in control mice injected with either AAV-control or AAV-LHPP (Extended Data Fig. 5c). Thus, the expression of LHPP prevents tumour formation and maintains liver function in L-dKO mice, suggesting that the protein histidine phosphatase LHPP is a tumour suppressor.

We next examined NME1, NME2 and LHPP expression in human liver tumours, from patients with HCC. Resected HCC tissue and adjacent non-tumour tissue were assessed for LHPP, NME1 and NME2 levels by immunoblot analysis. LHPP was low in the tumour (HCC) compared to non-tumour tissue in several patients (Fig. 4a), as observed in L-dKO mice. NME1 and NME2 expression was high in the tumours. We note that human NME1 and NME2, unlike their mouse counterparts, do not co-migrate (Fig. 4a). In agreement with the immunoblot analysis, immunohistochemistry also showed that LHPP expression was low and NME1 and NME2 expression was high in HCC compared to adjacent non-tumour tissue from two additional patients (Fig. 4b). Furthermore, using a tissue microarray assay, we evaluated LHPP expression in tumour and matched non-tumour liver samples from 20 patients with HCC. Tumour sections that had greater than 70% tumour cells and non-tumour sections that contained greater than 70% normal cells were chosen for the array (see Methods). LHPP expression was significantly downregulated in HCC compared to matched non-tumour tissue (Extended Data Fig. 6a–c). Similarly, on the basis

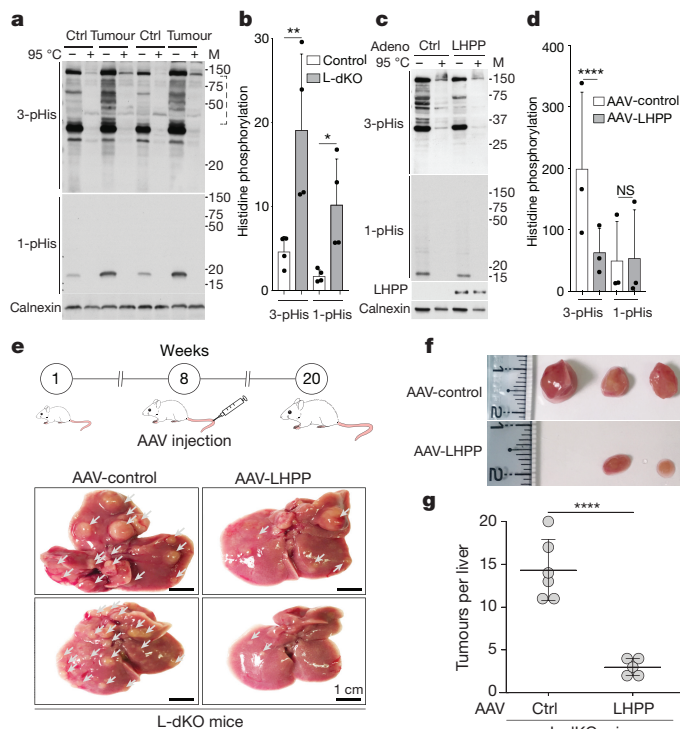


Figure 3 | LHPP is a protein histidine phosphatase and a tumour suppressor.

a, b, Immunoblot analysis (**a**) and quantification (**b**), shows increased 3-pHis and 1-pHis signals in L-dKO tumours (20 weeks) compared to control (ctrl) littermates. Similar results were observed in tumours from L-dKO mice ($n = 4$). Tissue lysates in $2\times$ sample buffer (pH 8.8) heated at 95°C for 10 min were dephosphorylated and served as a control. For full scans, see Supplementary Fig. 5. In the 3-pHis blot, the area enclosed in the dashed line was used for quantification ($n = 4$). Adjusted $P = 0.0460$, $**P = 0.0051$, two-sided, Holm–Sidak’s multiple comparisons test. Data are mean \pm s.d. M, molecular mass. **c**, Immunoblotting shows a reduction in 3-pHis levels in CB1 cells infected with adenovirus (adeno) containing LHPP, compared to cells infected with a control adenovirus. Similar results were observed in three biological experiments. For full scans, see Supplementary Fig. 7. **d**, Quantification of immunoblots indicate that 3-pHis signals are significantly reduced upon LHPP re-expression compared to control-infected cells ($n = 3$ biological repeats). $****P < 0.0001$, two-sided ratio paired t -test. NS, not significant. Data are mean \pm s.d. **e**, Top, experimental timeline. AAV-control or AAV-LHPP was injected into the tail vein of 8-week-old L-dKO and littermate control mice. Bottom, representative images of whole livers showing reduction in L-dKO liver tumour burden after infection with AAV-LHPP compared to infection with AAV-control (arrowheads indicate tumours). Similar results were observed in five different L-dKO mice infected with AAV-LHPP. **f**, Photographs of tumours isolated from L-dKO mice infected with AAV-LHPP and AAV-control. **g**, Quantification of tumours in L-dKO mice showing a significant reduction in tumour burden after infection with AAV-LHPP ($n = 5$ mice), compared to AAV-control ($n = 6$ mice). $****P < 0.0001$, two-sided unpaired t -test. Data are mean \pm s.d.

of a published mRNA expression dataset from 37 patients with HCC²¹, *LHPP* mRNA levels (but not *PHPT1* or *PGAM5* mRNA levels) were significantly reduced in tumour tissue compared to matched non-tumour tissue (Extended Data Fig. 6d). We note that, in L-dKO mice, reduced LHPP protein expression also correlated with reduced *Lhpp* mRNA levels (Extended Data Fig. 1f). Finally, consistent with increased levels of NME1 and NME2 and decreased levels of LHPP (Fig. 4c), we observed significantly higher 1-pHis and 3-pHis signals in human HCC compared to matched non-tumour tissue (Fig. 4c, d). Collectively, the above findings suggest that histidine phosphorylation is upregulated in both mouse and human tumours.

We next investigated clinical correlates of LHPP expression. HCC is scored based on Edmondson grades I to IV in which grade IV is

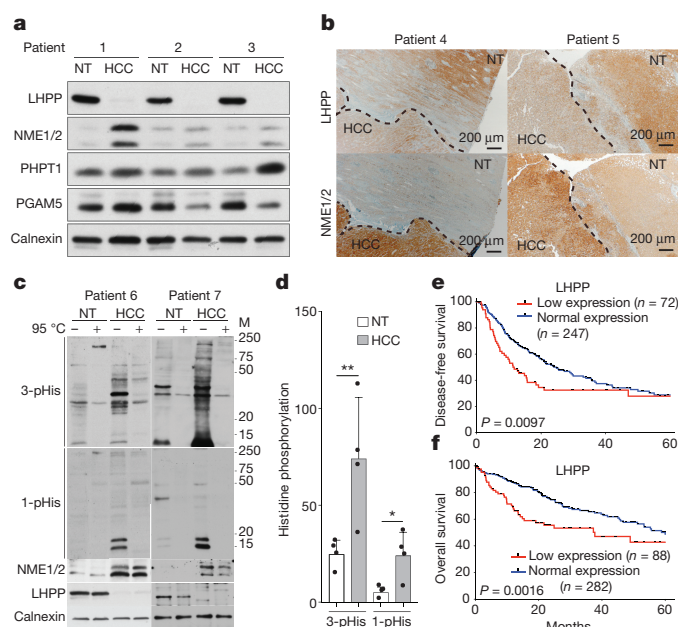


Figure 4 | Loss of LHPP expression in human HCC correlates with reduced patient survival. **a**, **b**, Immunoblot ($n = 3$ patients) (**a**) and immunohistochemistry (**b**) ($n = 2$ patients) analyses showing reduced LHPP and increased NME1 and NME2 expression in liver tissue from patients with HCC compared to adjacent non-tumour liver tissue in a total of $n = 7$ HCC patients ($n = 2$ other patients in panel **c**). For full scans, see Supplementary Fig. 12. **c**, Immunoblot analysis showing increased 3-pHis and 1-pHis signals in human HCC compared to adjacent non-tumour tissue. Similar results were observed in tumours from two additional HCC patients (total $n = 4$ patients). For full scans, see Supplementary Fig. 13. Tissue lysates in $2\times$ sample buffer (pH 8.8) heated at 95°C for 10 min to dephosphorylate pHis served as control. The monoclonal antibodies used to detect 3-pHis and 1-pHis are SC44-1 and SC1-1, respectively. **d**, Quantification of immunoblot (from **c** and two other western blots, $n = 4$ in total) indicates that 3-pHis and 1-pHis signals are significantly higher in human HCC compared to adjacent non-tumoural tissue (band intensities in each lane are normalized to the intensity of corresponding calnexin protein levels after subtracting the intensities from corresponding heated lane). The y axis represents band intensities. $*P = 0.0368$, $**P = 0.0064$, two-sided ratio paired t -test. Data are mean \pm s.d. **e**, **f**, Kaplan-Meier curves, log-rank test, showing the correlation between LHPP expression and clinical outcome, as analysed for disease-free survival (**e**) and overall survival (**f**).

the most severe or aggressive. Examination of a previously published mRNA dataset²¹ revealed that low LHPP expression is associated with more severe tumours (Edmondson grades III/IV versus I/II) (Extended Data Fig. 6e). These results, combined with our earlier observation that LHPP expression in L-dKO mice reduced tumorigenicity (Fig. 3e–g), suggest that low LHPP expression could correlate with reduced patient survival. We referred to The Cancer Genome Atlas (TCGA) to determine whether there is a correlation between LHPP expression and overall survival or disease-free survival (see Methods). LHPP downregulation was associated with faster disease progression and worse overall survival in 319 and 370 patients with HCC, respectively (Fig. 4e, f). Median disease-free survival and overall survival of patients with HCC decreased from 25 to 12 months and from 61 to 38 months, respectively, after the loss of LHPP expression. Thus, decreased LHPP expression in human HCC is associated with increased histidine phosphorylation and reduced survival.

We identified 49 LHPP mutations in datasets from TCGA and the International Cancer Genome Consortium (ICGC). Twelve (24.5%) of the mutations, found in oesophagus (2), head and neck (2), stomach (2), bladder (1), breast (1), skin (1), liver (1), lung (1), and pancreas (1), are predicted to be inactivating mutations (Extended Data Fig. 6f),

that is, nonsense, frameshift or splice site mutations (Supplementary Table 5). The $>20\%$ prevalence of inactivating mutations suggests that LHPP is a tumour suppressor gene²². This suggestion is strengthened by two genome-wide association studies (GWAS) indicating that the LHPP locus is associated with oral cancer, pharyngeal cancer and acute lymphoblastic leukaemia^{23,24}. Whole-genome sequencing previously identified the LHPP locus as a risk factor for major depressive disorder^{25,26}. These observations suggest that loss of LHPP and thus increased histidine phosphorylation is oncogenic and possibly also involved in other disorders. Importantly, the results collectively suggest that LHPP is a histidine phosphatase and tumour suppressor.

How might histidine phosphorylation be related to cancer? The histidine-phosphorylated proteins we detected preferentially in tumours include the crucial DNA replication factors MCM3, MCM6 and MCM7, the p53 inhibitor AROS, the cytidine deaminase ABEC3 that converts cytosine to uracil and may induce mutations, the receptor-mediated endocytosis and NOTCH signalling protein AAK1, the mTORC1 effector RSK2²⁷, and CHTOP. The potential LHPP targets we detected that might be important for tumorigenesis include the essential fatty acid biosynthetic enzyme ACLY⁶, the proto-oncoprotein and RNA polymerase-associated protein LEO1²⁸, a variant histone H2A commonly expressed in HCC²⁹, and the inflammation-related and HCC-promoting HMGB1 protein³⁰ (Supplementary Table 4).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 March 2017; accepted 14 February 2018.

Published online 21 March 2018.

- Kee, J.-M. & Muir, T. W. Chasing phosphohistidine, an elusive sibling in the phosphoamino acid family. *ACS Chem. Biol.* **7**, 44–51 (2012).
- Fuhs, S. R. *et al.* Monoclonal 1- and 3-phosphohistidine antibodies: new tools to study histidine phosphorylation. *Cell* **162**, 198–210 (2015).
- Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* **2**, 16018–16023 (2016).
- Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
- Kenerson, H. L. *et al.* Akt and mTORC1 have different roles during liver tumorigenesis in mice. *Gastroenterology* **144**, 1055–1065 (2013).
- Guri, Y. *et al.* mTORC2 promotes tumorigenesis via lipid synthesis. *Cancer Cell* **32**, 807–823 (2017).
- Yokoi, F., Hiraishi, H. & Izuhara, K. Molecular cloning of a cDNA for the human phospholysine phosphohistidine inorganic pyrophosphate phosphatase. *J. Biochem.* **133**, 607–614 (2003).
- Hiraishi, H., Yokoi, F. & Kumon, A. 3-phosphohistidine and 6-phospholysine are substrates of a 56-kDa inorganic pyrophosphatase from bovine liver. *Arch. Biochem. Biophys.* **349**, 381–387 (1998).
- Hiraishi, H., Ohmagari, T., Otsuka, Y., Yokoi, F. & Kumon, A. Purification and characterization of hepatic inorganic pyrophosphatase hydrolyzing imidodiphosphate. *Arch. Biochem. Biophys.* **341**, 153–159 (1997).
- Ek, P. *et al.* Identification and characterization of a mammalian 14-kDa phosphohistidine phosphatase. *Eur. J. Biochem.* **269**, 5016–5023 (2002).
- Klumpp, S. *et al.* Protein histidine phosphatase: a novel enzyme with potency for neuronal signaling. *J. Cereb. Blood Flow Metab.* **22**, 1420–1424 (2002).
- Panda, S. *et al.* Identification of PGAM5 as a mammalian protein histidine phosphatase that plays a central role to negatively regulate CD4⁺ T Cells. *Mol. Cell* **63**, 457–469 (2016).
- Cai, X., Srivastava, S., Surindran, S., Li, Z. & Skolnik, E. Y. Regulation of the epithelial Ca²⁺ channel TRPV5 by reversible histidine phosphorylation mediated by NDPK-B and PHPT1. *Mol. Biol. Cell* **25**, 1244–1250 (2014).
- Hartsough, M. T. *et al.* Nm23-H1 metastasis suppressor phosphorylation of kinase suppressor of Ras via a histidine protein kinase pathway. *J. Biol. Chem.* **277**, 32389–32399 (2002).
- Wagner, P. D. & Vu, N. D. Phosphorylation of ATP-citrate lyase by nucleoside diphosphate kinase. *J. Biol. Chem.* **270**, 21758–21764 (1995).
- Fuhs, S. R. & Hunter, T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Curr. Opin. Cell Biol.* **45**, 8–16 (2017).
- Kee, J.-M., Villani, B., Carpenter, L. R. & Muir, T. W. Development of stable phosphohistidine analogues. *J. Am. Chem. Soc.* **132**, 14327–14329 (2010).
- Riley, N. M. & Coon, J. J. Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.* **88**, 74–94 (2016).
- Attwood, P. V., Piggott, M. J., Zu, X. L. & Besant, P. G. Focus on phosphohistidine. *Amino Acids* **32**, 145–156 (2007).

20. Besant, P. G. & Attwood, P. V. Detection and analysis of protein histidine phosphorylation. *Mol. Cell. Biochem.* **329**, 93–106 (2009).
21. Makowska, Z. *et al.* Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma. *J. Pathol. Clin. Res.* **2**, 80–92 (2016).
22. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
23. Vijayakrishnan, J. *et al.* A genome-wide association study identifies risk loci for childhood acute lymphoblastic leukemia at 10q26.13 and 12q23.1. *Leukemia* **31**, 573–579 (2017).
24. Lesseur, C. *et al.* Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat. Genet.* **48**, 1544–1550 (2016).
25. Neff, C. D. *et al.* Evidence for HTR1A and LHPP as interacting genetic risk factors in major depression. *Mol. Psychiatry* **14**, 621–630 (2009).
26. CONVERGE consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
27. Park, I.-H., Bachmann, R., Shirazi, H. & Chen, J. Regulation of ribosomal S6 kinase 2 by mammalian target of rapamycin. *J. Biol. Chem.* **277**, 31423–31429 (2002).
28. Chong, P. S. Y. *et al.* LEO1 is regulated by PRL-3 and mediates its oncogenic properties in acute myelogenous leukemia. *Cancer Res.* **74**, 3043–3053 (2014).
29. Lo, Re, O. *et al.* Induction of cancer cell stemness by depletion of macrohistone H2A1 in hepatocellular carcinoma. *Hepatology* **67**, <http://doi.org/10.1002/hep.29519> (2018).
30. Chen, M. *et al.* High-Mobility Group Box 1 promotes hepatocellular carcinoma progression through miR-21-mediated matrix metalloproteinase activity. *Cancer Res.* **75**, 1645–1656 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.N.H. acknowledges the Louis Jeantet Foundation, the Swiss National Science Foundation, SystemsX.ch, and the European Research Council (MERiC). T.H. acknowledges USPHS grants CA080100, CA082683 and CA194584 from the NCI. T.H. is an American Cancer Society Professor, and holds the Renato Dulbecco Chair in Cancer Research. S.P. acknowledges the Swiss National Science Foundation (Ambizione grant number PZ00P3_168165).

Author Contributions S.K.H. and M.N.H. conceived and designed the experiments, and wrote the manuscript. M.Col. maintained the in-house proteome database. S.R.F., K.A. and T.H. prepared pHis antibodies. L.M.T., L.Q. and M.S.M. performed histological analysis on patient tissues. D.L., C.B. and Y.G. assisted with animal experimentation. M.Cor. generated the L-dKO mouse. S.P. and C.K.Y.N. analysed public databases for mutations, gene expression and patient survival. S.M. and P.J. assisted with mass spectrometry. M.H.H. provided HCC microarray data. All authors commented and agree on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.N.H. (m.hall@unibas.ch).

METHODS

Animals. Liver-specific *Tsc1* and *Pten* double-knockout mice were obtained by crossing *Tsc1*^{lox/lox} mice (exons 17 and 18)³¹ with *Pten*^{lox/lox} mice (exons 4 and 5)³² to transgenic mice expressing Cre recombinase under the control of the hepatocyte-specific albumin promoter (Alb-CreTg/0)³³, to generate liver-specific double-knockout (*Tsc1*^{lox/lox}*Pten*^{lox/lox} Alb-CreTg/0, in short L-dKO) mice. Age- and sex-matched littermate *Tsc1*^{lox/lox}*Pten*^{lox/lox} mice without the *cre* genes were used as controls in all animal experiments. No other randomization protocol or blinding was applied beyond this criterion. The mice produced were on mixed genetic background (C57BL/6J, 129/SvJae, BALB/c). Mice were housed under temperature and humidity-controlled conditions, in a 12-h light/dark cycle with lights switched on between 06:00 to 18:00. All experiments were conducted on male mice (age between 6 and 20 weeks). In all experiments, mice were fasted overnight before euthanasia by CO₂ inhalation. All relevant ethical regulations were followed in this study. The tumour harvest time limit was set at 20 weeks. In addition, the clinical conditions of mice were monitored according to the score sheet as stipulated by the Institutional Animal Care and Use Committee (IACUC), Kantonales Veterinäramt of Kanton Basel-Stadt. None of the guidelines was exceeded in any experiment performed. For all animal experiments, respecting the animal ethical rules, a minimum number of animals required to get a statistically meaningful result were used. All animal experiments were performed in accordance with the federal ethical guidelines and were approved by the Kantonales Veterinäramt of Kanton Basel-Stadt.

Mass spectrometry. Sample preparation for proteome analysis of tumours. Tumours from 20-week-old L-dKO mice and whole liver from littermate control mice were carefully dissected under magnifying lens and snap-frozen in liquid nitrogen. Frozen tissue was pulverized in a metal plate cooled on dry ice and was transferred into a cooled microcentrifuge tube. The powdered tissue was resuspended in 8 M urea (Applichem, A1086) (containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM PMSE, 1 × Complete Mini Protease Inhibitors (Roche), 1 × PhosSTOP (Roche)) and homogenized using a Polytron (PT 10-35 GT) at 500g for 2 min. The sample was placed on ice during this process. Lysates were then subjected to rotation at 4°C with regular vortexing. Lysates were then centrifuged at 10,000g for 10 min at 4°C to remove cell debris. Protein concentration in the supernatant was determined using Bradford assay. Equal amounts (5 mg each) of 'pooled control' and tumour tissue were reduced with 2.5 mM dithiothreitol (DTT) (Sigma) for 45 min at 56°C, cooled for 5–10 min at room temperature and alkylated with 7.5 mM iodoacetamide (IAA) (Sigma) for 45 min at room temperature (reaction protected from light). The urea concentration was reduced to 4 M by diluting with 25 mM Tris-HCl, pH 8.0. Proteins were treated with endoproteinase LysC (1:100 (w/w), Wako) at 37°C for 2 h. The urea concentration of the LysC digest was reduced to 1 M. Later, trypsin (Worthington) digestion (1:100; w/w) was performed for 2 h at 37°C, followed by a second round of trypsin digestion (1:100; w/w) overnight at 37°C. Digestion was stopped by adding 0.4% trifluoroacetic acid (TFA) (Applied Biosystems) (v/v) that lowered the pH of the solution to below pH 2.0. Later, the digest was centrifuged at 10,000g for 10 min at room temperature and the supernatant was desalted on a C18 reverse-phase SepPak-200mg column (Waters). First, the cartridge was primed with 2 ml 100% acetonitrile (ACN) (Biosolve), followed by 4 ml of 50% acetonitrile containing 0.5% acetic acid (Merck) and equilibrated with 4 ml 0.1% TFA. Peptides were loaded twice in 0.4% TFA and the cartridge was washed with 10 ml of 0.1% TFA. Later, the cartridge was washed with 0.5% acetic acid and the peptides were eluted with 4 ml of 80% ACN, 0.5% acetic acid. The peptides were dried in a SpeedVac and subjected to strong cation exchange chromatography (SCX). SCX fractionation was performed according to³⁴ with a few modifications detailed below. Storage buffer was removed from a HiTrap SP cartridge (GE Healthcare, 17115101) and washed twice with 2 ml of SCX buffer A (5 mM KH₂PO₄ (pH 2.65), 30% can). Later, the cartridge was equilibrated twice with 2 ml of SCX buffer B (5 mM KH₂PO₄ (pH 2.65), 30% ACN containing 500 mM KCl), and the cartridge was re-equilibrated with 2 ml of SCX buffer A. The peptides were resuspended in 1 ml of SCX buffer A, centrifuged at 10,000g for 2 min to remove any unwanted debris. The peptides were then loaded with a syringe onto the HiTrap SP cartridge. The flow-through was then collected. The bound peptides were then desorbed in a stepwise manner: with 1 ml each of SCX buffer A containing 50 mM, 100 mM, 150 mM, 250 mM and 350 mM KCl. Each fraction was collected and the peptide concentration was estimated at 280 nm. The fractions were dried in a SpeedVac and peptides were then desalted on Microspin cartridges (Vydac). 10 µg of each fraction was removed for liquid chromatography–tandem mass spectrometry (LC–MS/MS proteome analysis).

LC–MS/MS analysis. Dried peptides were dissolved in 20 µl of 0.1% acetic acid, 0.005% TFA. The peptides were analysed by capillary LC–MS/MS using a home-made separating column (0.075 mm × 18 cm) packed with Reprosil C18 reverse-phase material (2.4 µm particle size). The column was connected on line to an Orbitrap FT hybrid instrument (Thermo Scientific). Solvent A (0.1% acetic acid

in water, 0.005% TFA) and solvent B (0.1% acetic acid, 0.005% TFA and 80% acetonitrile in water) were used for peptide separation. 2 µl of peptide digest were injected with a Proxeon Easy-LC capillary pump (Thermo Scientific) set to 0.3 µl min⁻¹. A gradient from 0 to 40% solvent B in solvent A in 190 min was delivered with the nano pump at a flow rate of 300 nL min⁻¹. The percentage of solvent B was increased to 75% in 10 min at the end of the gradient. Later, the eluting peptides were ionized at 2.5 kV and mass spectrometer was operated in data-dependent mode. The precursor scan was done in the Orbitrap set to a 60,000 resolution, while the fragment ions were mass analysed in the LTQ instrument. A top twenty method was used for fragmentation.

Protein identification. The LC–MS/MS data were searched with Proteome Discoverer 1.4 (Thermo Scientific) set to Mascot and Sequest HT search engines with 10 ppm precursor ion tolerance. The fragment ions were set to 0.6 Da tolerance. Medium confidence (5% FDR) was used for peptide search. Swiss-Prot KB database set to *M. musculus* was used for Mascot. For Sequest HT, an inhouse databank was constructed by extracting *M. musculus* entries from UniProtKB.

Quantification. The raw files from the LC–MS/MS were first passed through Progenesis Q1 software (Nonlinear Dynamics, Waters) and the matched SCX fractions between 'pooled control' and tumours were aligned. A minimum of 75% alignment was set for analysis. During alignment, the search identifications from Proteome Discoverer (version 1.4) were imported into Progenesis for matching the aligned features with the identified peptides.

Statistical analysis. R-based Perseus program, version 1.4.0.2, was used for statistical data analysis. An ANOVA-based two-sample *t*-test was performed, adjusting S0 to 1, the number of randomizations to 250 (default), and the FDR to 5% for the proteome^{35,36}. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001.

p-His antibody immunoprecipitation of proteins for mass spectrometry. Tissue and cell lysis for pHis immunoprecipitation. Tumours from 20-week-old L-dKO mice and whole liver from littermate control mice were carefully dissected under magnifying lens and snap-frozen in liquid nitrogen. Frozen tissue was pulverized in a metal plate cooled on dry ice and was transferred into a cooled microcentrifuge tube. The powdered tissue was resuspended in 25 mM Tris-HCl (pH 8.5); 100 mM NaCl, pH: 8.5; 1 mM PMSE, 1 × Complete Mini Protease Inhibitors (Roche), 1 × PhosSTOP (Roche) and homogenized using a Polytron (PT 10-35 GT) at 500g for 2 min. Precautions were taken to preserve pHis (pH 8.5 and constant 4°C) and reagents (such as detergents) that can interfere with mass spectrometry. The sample was placed on ice during this process. Lysates were then incubated at 4°C on an end-over-end rotator for 2 h with regular vortexing. Lysates were then centrifuged at 10,000g for 10 min at 4°C to remove cell debris. Protein concentration in the supernatant was determined using BCA assay. Tissue lysate from four different littermate control animals was pooled to obtain a pooled control. For immunoprecipitation of pHis proteins from CB1 cells with and without LHPP, 2 mg for each condition was used.

Bead preparation and preclearing of the lysate. One millilitre of protein A Sepharose beads (Thermo, 101042) was mixed with 1 ml of lysis buffer (25 mM Tris-HCl (pH 8.5); 100 mM NaCl, pH: 8.5; 1 mM PMSE, 1 × Complete Mini Protease Inhibitors (Roche), 1 × PhosSTOP (Roche)) and washed twice (by spinning at 2,000g, 4 min, 4°C) and resuspended in 2 ml of lysis buffer. For pre-clearing, the protein lysate (10 mg for tissues, 2 mg for cells) was incubated with 200 µl of prepared beads and incubated at 4°C on an end-over-end rotator for 2 h. Later, the lysate was spun (2,000g, 4 min, 4°C) and the precleared lysate (supernatant) was divided into two different tubes (5 mg each for tissues, 1 mg each for cells).

Antibody-lysate incubation. For tissue lysates, both 1-pHis and 3-pHis (clones SC1-1 and SC44-1) antibodies were used. For cell lysates, only 3-pHis antibody was used. Lysates were incubated with pHis antibody (30 µg of 3-pHis plus 10 µg of 1-pHis for tissue lysate; 30 µg of 3-pHis for cell lysate) or IgG antibody (40 µg for tissue lysate and 30 µg for cell lysate). Lysates were incubated at 4°C on an end-over-end rotator for 2 h. Later, 200 µl of beads in lysis buffer were added and incubation continued at 4°C on an end-over-end rotator for 12 h.

Washing the beads and elution of the proteins. The column was washed four times with 2 ml of lysis buffer (pre-cooled) at 4°C. After the third wash, the fourth wash was performed in a new microcentrifuge tube to avoid any contamination from proteins attached to the wall of the microcentrifuge tube. Proteins were eluted in two fractions (E1 and E2) with 2 × 600 µl 100 mM triethylamine (TEA), pH 11. The fractions were dried in a SpeedVac and resuspended in 50 mM Tris-HCl (pH 7.5).

Sample preparation. The eluate in 50 mM Tris-HCl was treated with endoproteinase LysC (1:100 (w/w), Wako) at 37°C for 2 h. Later, trypsin (Sigma) digestion (1:100; w/w) was performed for 2 h at 37°C (pH 8.0), followed by a second round of trypsin digestion (1:100; w/w) overnight at 37°C. Digestion was stopped by adding 0.4% TFA (Applied Biosystems) (v/v) that lowered the pH of the solution to below pH 2.0. Later, the digest was centrifuged at 10,000g for 10 min at room temperature and the supernatant was desalted on Microspin cartridges (Vydac).

First, the cartridge was activated with 100 µl methanol, followed by 100 µl 80% ACN, 0.5% acetic acid and the column was equilibrated with $2 \times 100 \mu\text{l}$ 0.1% TFA, 1% ACN. Peptides were loaded twice in 0.4% TFA and the cartridge was washed five times with 100 µl 0.1% TFA, 1% ACN. Later, the peptides were eluted with $2 \times 100 \mu\text{l}$ 80% ACN, 0.5% acetic acid. The peptides were dried in a SpeedVac and subjected to LC-MS/MS proteome analysis. For further details, please see the 'LC-MS/MS analysis' section above.

Protein identification. The LC-MS/MS data were searched with MaxQuant³⁷, version 1.5.3.8 and searched against the mouse Swiss-Prot database 2017-05-09. A mass tolerance of 20 p.p.m. was allowed for the first search, and a mass tolerance of 10 p.p.m. was allowed for the main search. Two missed cleavages were allowed. Peptide FDR, protein FDR, and site decoy fraction were set to 2%. For protein quantification, a ratio count of 1 was applied. The fragment ions were set to 0.6 Da tolerance. Medium confidence (5% FDR) was used for peptide search.

Quantification. The protein groups datasets were exported into a FileMaker Pro-12 databank.

Identification of proteins with histidine phosphorylation in tumours. Samples were run in triplicates. Label-free quantification (LFQ) intensities were obtained for each protein from MaxQuant. A mean of LFQ intensities was calculated and used for all further calculations. Proteins without any ratio owing to no detection in either of the samples were also included in the analysis. Mass spectrometry analysis identified 3,265 proteins in the tumour immunoprecipitate from a total of 4 different L-dKO tumours. A total of 1,508 proteins were at least twofold-enriched in the pHis immunoprecipitate compared to the control IgG immunoprecipitate in more than 1 tumour. Out of these, 1,173 proteins were detected and enriched (minimum twofold) in 2 out of 2 tumours (100%), or 3 out of 3 tumours (100%), or 3 out of 4 tumours (75%, 100%) (Supplementary Table 2).

Identification of proteins with histidine phosphorylation in control tissue. Mass spectrometry analysis identified 3,011 proteins in the control tissue immunoprecipitate from a total of 4 different control animals. A total of 1,141 proteins were at least twofold-enriched in the pHis immunoprecipitate compared to the control IgG immunoprecipitate in more than 1 tumour. Out of these, 823 proteins were detected and enriched (minimum twofold) in 2 out of 2 control tissues (100%), or 3 out of 3 control tissues (100%), or 3 out of 4 control tissues (75%, 100%) (Supplementary Table 1).

Identification of proteins with increased histidine phosphorylation in tumours. Out of the proteins detected in tumour (1,173) and control (823) tissue, 236 proteins were detected and enriched (minimum twofold) in 2 out of 2 experiments (100%), or 3 out of 3 experiments (100%), or 3 out of 4 experiments (75%, 100%) (Supplementary Table 3). A total of 24 proteins that are exclusively detected in tumour tissue and hence had no ratio are indicated in the list shown in Supplementary Table 3.

Identification of proteins with decreased histidine phosphorylation in CB1 cells upon LHPP re-expression. Mass spectrometry analysis identified 295 proteins in the CB1 immunoprecipitate from a total of 5 different experiments. Nine proteins were at least 1.7-fold enriched in the CB1 cells lacking LHPP compared to the control IgG immunoprecipitate and LHPP re-expressing cells in minimum 50% of the experiments (Supplementary Table 4).

Antibodies. PTEN (CST, 9188), TSC1 (Bethyl, A300-316A), phospho-S6 (CST, 4856), total S6 (CST, 2217), pAKT S473 (CST, 9271), total AKT (CST, 9272), Calnexin (Enzo, ADI-SPA-860), NME1 (CST, 3345), NME2 (Abcam, ab60602), and LHPP (Novus, NBPI-83273) were obtained commercially. Phospho-histidine monoclonal antibodies against N1 and N3 positions were as described previously². The monoclonal antibodies used to detect 3-pHis and 1-pHis were SC44-1 and SC1-1, respectively. Horseradish peroxidase (HRP)-coupled anti-mouse (115-035-774) and anti-rabbit (211-032-171) secondary antibodies were purchased from Jackson laboratories.

Immunoblot analysis. For regular immunoblot analysis, the liver tissue was homogenized in T-PER (ThermoFisher scientific, 78510) supplemented with 1 mM PMSF, 1 × Complete Mini Protease Inhibitors (Roche), 1 × PhosSTOP (Roche) using a Polytron (PT 10-35 GT) at 500g for 2 min. Equal amounts of homogenate were SDS-PAGE fractionated and transferred onto a nitrocellulose membrane that was incubated, after blocking (5% BSA in TBST), with appropriate antibodies. For phosphohistidine analysis, tissues (equal wet weights) and equal number of cells were lysed directly in 2 × pHis sample buffer (5 × pHis sample buffer: 250 mM Tris-HCl pH 8.8, 0.02% bromophenol blue, 50% glycerol, 50 mM EDTA, 500 mM DTT and 10% SDS), incubated on ice and sonicated (15 s × 3 times, Hielscher VialTweeter) to disrupt cells and shear DNA. Lysates were clarified by centrifugation (10,000g for 15 min at 4 °C). Fresh lysates were resolved immediately using polyacrylamide gels at 4 °C with a modified stacking gel, pH 8.8 and 10% resolving gel, pH 8.8. All electrophoresis and protein transfer was strictly carried out at 4 °C. In addition, the running and transfer buffers were precooled before use.

The nitrocellulose membranes were blocked in casein blocking buffer (0.1% casein, 10% PBS, pH 8.8). Primary antibodies were resuspended in casein blocking buffer with 0.1% Tween-20, incubated overnight at 4 °C. The monoclonal antibodies used to detect 3-pHis and 1-pHis were SC44-1 and SC1-1, respectively at a dilution of 0.5 µg ml⁻¹. The membranes were washed three times (15 min per wash at 4 °C) with TBST (modification, pH 8.8) and incubated with anti-rabbit HRP conjugated secondary antibody (1:10,000) at 4 °C. Lysates in sample buffer (pH 8.8) heated at 95 °C for 10 min to dephosphorylate pHis served as a control. The bands in the heated lane are either heat-resistant pHis bands or a non-specific signal. For quantification of the pHis immunoblots, band intensities in each lane are normalized to intensity of corresponding calnexin protein levels after subtracting the intensities from corresponding heated lane; the y-axis represents band intensities.

Immunoblot analysis on Lhpp adenovirus-infected cells. A 50–60% confluent dish was first infected, 24 h after infection, the cells (CB1 or SNU449) were trypsinized and 2×10^5 cells were seeded into a 6-well adherent tissue culture dish and 24 h later lysed in pHis compatible buffer (2 × pHis sample buffer, for buffer composition, see the 'Immunoblot analysis' section above).

Crystal violet staining. On Lhpp-adenovirus infected cells. Twelve hours after infection, CB1 (2×10^4 per well) and SNU449 (4×10^4 per well) cells were seeded into a 6-well plate and stained with crystal violet (2% crystal violet in 20% methanol) on day 7 ($n = 4$). The plate was left to dry at room temperature. Absorbance was measured at 560 nm on a Tecan Infinite M1000 machine. The software used to control was Tecan i-control, version 1.11.1.0, according to the manufacturer's instructions. Reference (or blank) was set with an empty 6-well plate. Multiple reads were read per well (15 × 15 reads) and a border of 3,000 µm was left in each well. Number of flashes was set to 25 with 10 ms of settle time.

On LHPP-siRNA-transfected cells. LHPP siRNA (human LHPP ON-TARGET plus set of 4, pooled, catalogue number LQ-018950-02-0005; siRNA 1 CAACCCAAACUGUGUGGUA, siRNA 2 CAUGAAGGCGCUUGAGUAU, siRNA 3 GCAGCAGCCGACAAGUGA, siRNA 4 CUGAAGCGUCCCCG GCUGA) and negative control siRNA (ON-TARGET plus non-targeting pool) were purchased from Dharmacon. SNU449 cells were transfected in the presence of the transfection reagent jetPRIME (PolyPlus transfection) as per the manufacturer's instructions. Cells were then re-suspended in complete DMEM plus 10% FBS medium. Then 36 h after transfection, SNU449 cells were re-transfected with LHPP siRNA and were seeded in a 6-well plate (4×10^4 per well). The cells were stained with crystal violet (2% crystal violet in 20% methanol) on day 3.

Cell culture. CB1 cells were established from 20-week-old L-dKO tumours. Multiple L-dKO tumours were chopped finely into small pieces and washed in warm 20 ml PBS. The processed tumour tissues were seeded in a 6-well tissue culture dish in DMEM with 10% FBS. Six different clones, CB1–CB6 were selected by limited trypsinization and amplified in DMEM with 10% FBS. Clone CB1 (transformed *in vivo*) was used for all future assays. SNU449 cells were obtained from ATCC and grown in DMEM with 10% FBS. We did not authenticate the cell line in our laboratory. The CB1 cell line was authenticated as hepatocytes by probing for the markers TSC1 and PTEN that were knocked out only in hepatocytes. CB1 and SNU449 cell lines were tested and found to be free of mycoplasma contamination.

In vitro phosphatase assay. CB1 cells were lysed in 2 × sample buffer (pH 8.8) and electrophoresed in two sets using a 10% modified SDS gel (see the 'Immunoblot analysis' section above). The proteins were transferred onto a nitrocellulose membrane. After transfer, the membrane was cut into two halves. One half of the membrane was treated with 0.5 µg of recombinant LHPP (Abcam, ab116175) in 500 µl Tris-MgCl₂ buffer (20 mM Tris-HCl, pH 8.0, and 5 mM MgCl₂). The other half of the membrane was treated with 1 ml Tris-MgCl₂ buffer. The assay was carried out at room temperature. After 4 h of treatment with LHPP, the blots were washed in casein blocking buffer (pH 8.8) and immunoblotted with 3-pHis antibody. The monoclonal antibody used to detect 3-pHis was SC44-1.

INK128 treatment and AAV administration. For acute INK128 administration, 20-week-old L-dKO ($n = 8$) and control animals ($n = 8$) were injected with INK128 (1 mg kg⁻¹ of body weight) or with the vehicle (intraperitoneal injection). INK128 was dissolved in 5% 1-methyl-2-pyrrolidone, 15% polyvinylpyrrolidone K30, and 80% water. INK128 injection was performed twice (06:00 and 18:00) over 24 h. For AAV administration, 8-week-old control and L-dKO mice were infected with AAV/DJ-ALBP-RFP (AAV-control) or AAV/DJ-ALBP-RFP-2A-mLhpp (AAV-LHPP) (5×10^{11} genome copies per mouse) via tail vein injection.

Hepatosphere assays. CB1 cells (2×10^5 cells per well) were seeded in 6-well ultra-low attachment plates (Corning, 3471, Costar 6 Well Plate with lid) in serum-free DMEM/F12 with recombinant murine epidermal growth factor (20 ng ml⁻¹; R&D Systems), murine hepatocyte growth factor (Sigma-Aldrich), and 1 × B27 serum-free supplement (Invitrogen). Cells were cultured at 37 °C, in 5% CO₂. On the fifth day, hepatospheres measuring greater than 100 µm were counted for the assay.

Histopathology and immunohistochemistry. After euthanization of mice, the livers were fixed in 4% paraformaldehyde, dehydrated and embedded into paraffin wax blocks. Embedded-tissues were cut into 4- μ m-thick sections placed on SuperFrost slides (Thermo Scientific) and stained with H&E (Merck). Immunohistochemistry was performed using the antibodies NME1 (CST, 3345, dilution 1:100), NME2 (Abcam, ab60602, 1:100), and LHPP (Novus, NBP1-83273, 1:200).

HCC tissue specimens and tissue microarrays. Formalin-fixed and paraffin-embedded tumour specimens, as well as fresh frozen tissue were obtained from the Institute of Pathology, University of Basel, Switzerland. For tissue microarray construction, a representative tumour area was selected on an H&E-stained slide of the donor block. Afterwards, a core punch with a diameter of 0.6 mm was taken from the tumour, and in selected cases from the non-tumoral liver tissue. Core punches were transferred to a new paraffin recipient block by using a programmed tissue arrayer (Beecher Instruments). The tissue microarray contained 30 HCC samples and 30 non-tumour liver control samples. Microarrays were cut in sections of 4 μ m thickness and stained with a polyclonal LHPP antibody diluted at 1:200 using an automatic Benchmark XT staining machine (Ventana Medical Systems Inc.). LHPP staining intensity was evaluated by a clinical pathologist (M.S.M.) and graded semiquantitatively into: 0 for negative staining, 1 for weak positive staining, 2 for moderate positive staining and 3 for strong positive staining (Extended Data Fig. 6b). Ten pairs of samples in which one of the tissue sections was washed away during staining were ignored during evaluation. While scoring for staining intensity, the pathologist was double blinded to the antigen that was probed and the sample identity.

Re-analysis of transcriptomic profiling data. LHPP expression was evaluated in 59 HCC needle biopsies, 37 of which were matched with their corresponding non-neoplastic liver parenchyma (cirrhotic tissues), using published transcriptomic data (GSE64041)²¹. CEL files were normalized using the Qlucore Omics Explorer software (Qlucore AB). LHPP expression was extracted for each sample. RNA sequencing gene expression data including outcomes from 370 hepatocellular carcinomas were retrieved from The Cancer Genome Atlas dataset (TCGA, Provisional) using the cBioportal (<http://www.cbioportal.org>, accessed 4 November 2016) website. Downregulation of LHPP was defined as z-score ≤ -1 .

LHPP mutation analysis. We interrogated the cBioPortal^{38,39} and ICGC data portal (<https://dcc.icgc.org>) for deletions at the LHPP locus in samples across

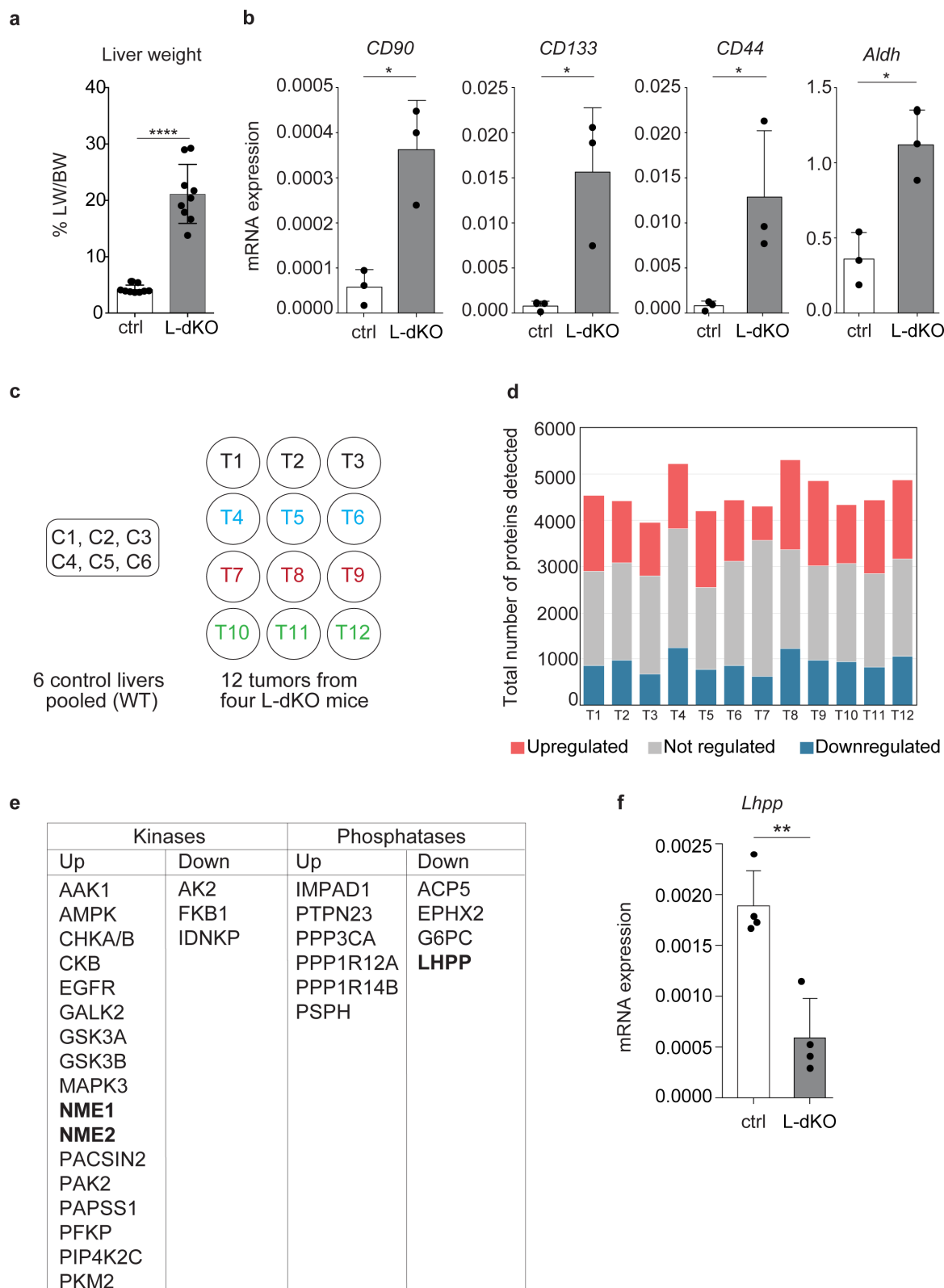
33 cancer types. We identified 32 cases harbouring deletion at the LHPP locus. Furthermore, we identified 49 mutations across the datasets. We removed duplicate TCGA samples that appear in both data portals and variants with >1% population frequencies in the Exome Aggregation Consortium data set, ExAC⁴⁰. 12 out of 49 (24.5%) of the mutations are inactivating mutations (Supplementary Table 5).

Serum analysis. Liver damage enzymes ALT, AST and LDH were measured using a biochemical analyser (Cobas c 111 analyser, Roche).

Patient material and ethics. All relevant ethical regulations were followed in this study. All the experiments with human tissue samples reported in this study were approved by the ethics commission of Northwestern Switzerland (EKNZ, approval No. 361/12). An informed consent was obtained from the human subjects.

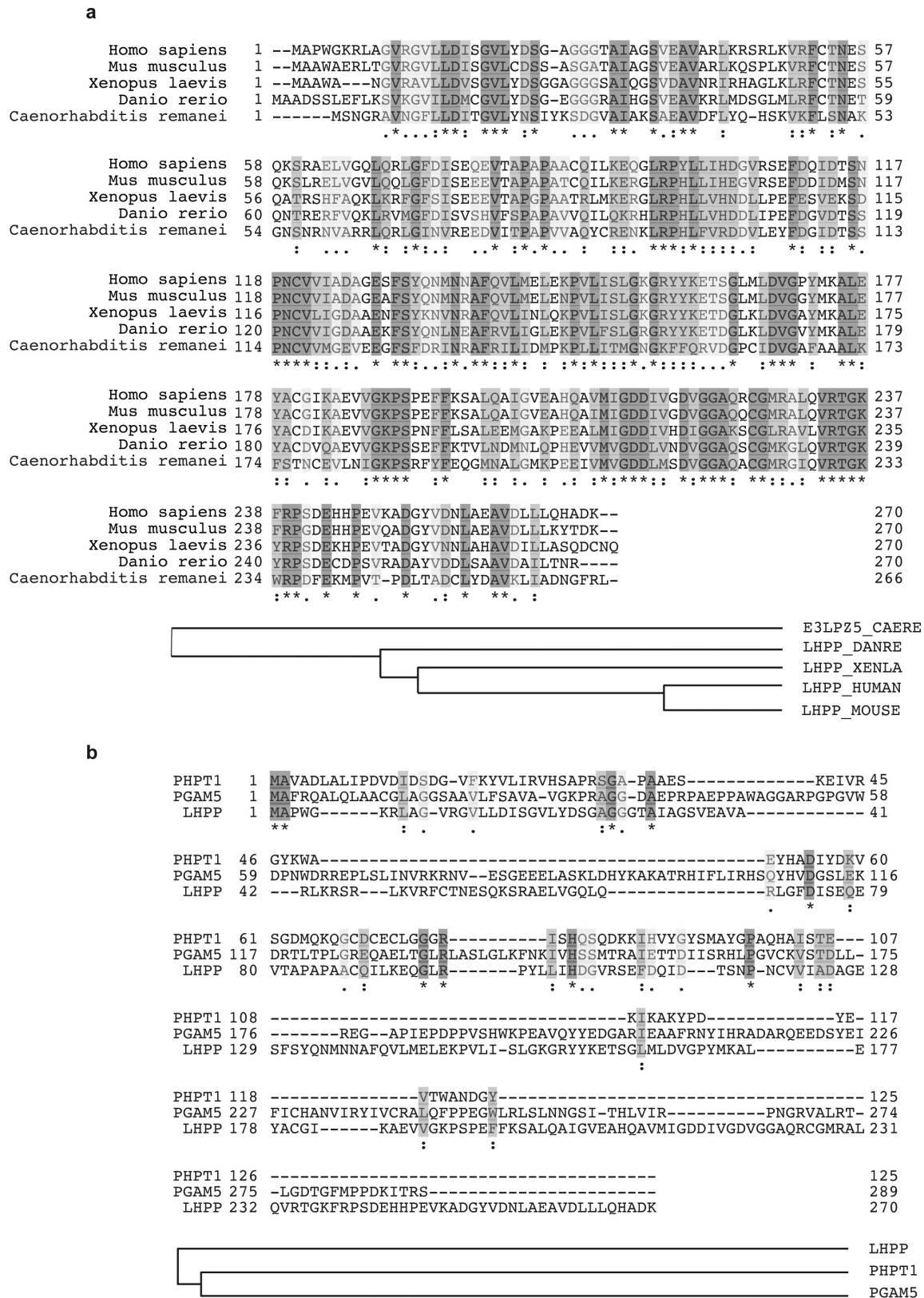
Data availability. All data generated during this study are included in this paper and its figures, Extended Data and Supplementary Information. For gel source data, see Supplementary Figs 1–13. In addition, publicly available data were analysed in this study and are cited wherever relevant.

31. Kwiatkowski, D. J. *et al.* A mouse model of TSC1 reveals sex-dependent lethality from liver hemangiomas, and up-regulation of p70S6 kinase activity in Tsc1 null cells. *Hum. Mol. Genet.* **11**, 525–534 (2002).
32. Horie, Y. *et al.* Hepatocyte-specific Pten deficiency results in steatohepatitis and hepatocellular carcinomas. *J. Clin. Invest.* **113**, 1774–1783 (2004).
33. Postic, C. & Magnuson, M. A. DNA excision in liver by an albumin-Cre transgene occurs progressively with age. *Genesis* **26**, 149–150 (2000).
34. Dephoure, N. & Gygi, S. P. A solid phase extraction-based platform for rapid phosphoproteomic analysis. *Methods* **54**, 379–386 (2011).
35. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
36. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
37. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
38. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
39. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
40. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

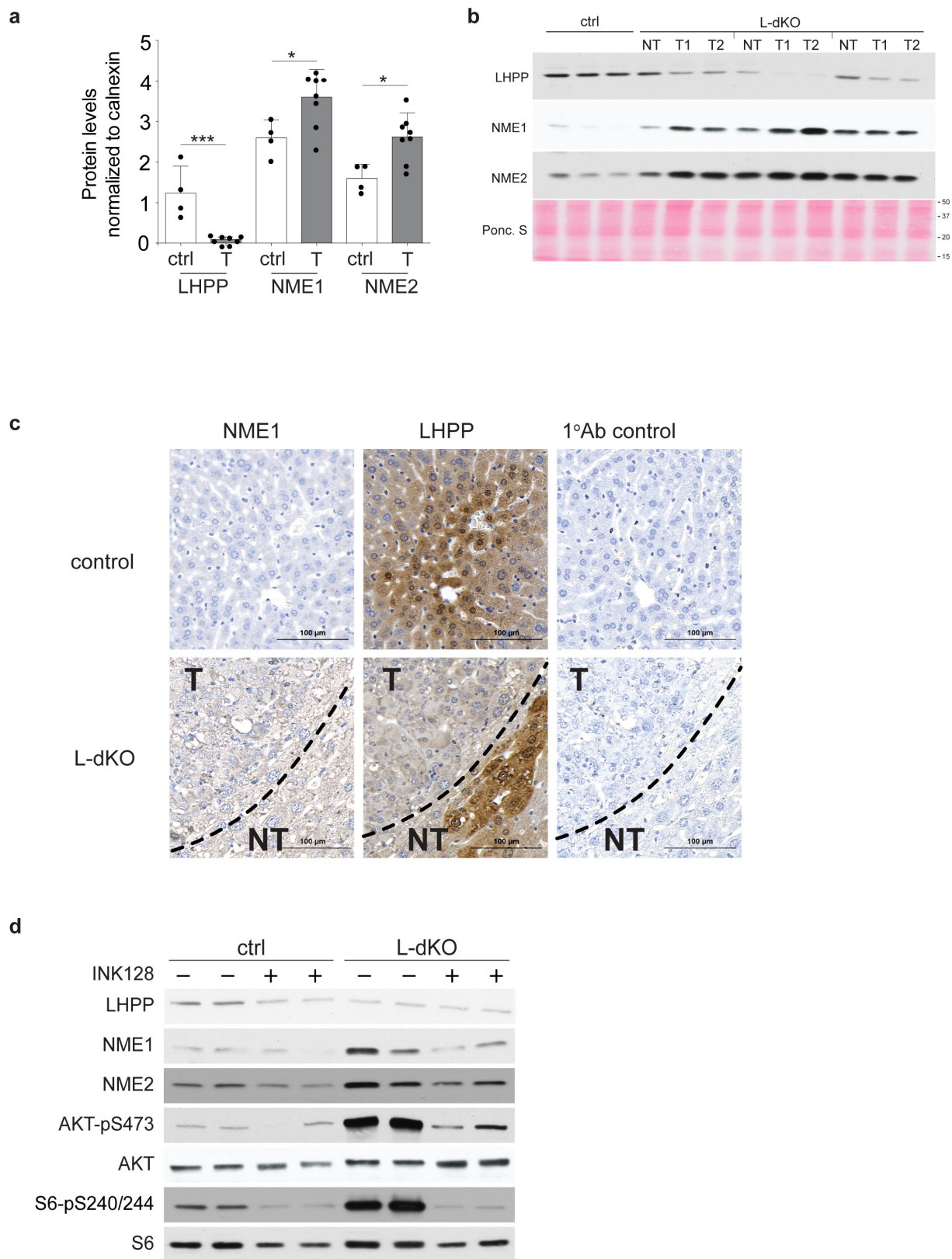


Extended Data Figure 1 | Proteomics on L-dKO tumours. **a**, Liver weight (LW) to body weight (BW) ratio (percentage) of 20-week-old L-dKO mice ($n=9$) and age-matched control mice ($n=9$). **** $P < 0.0001$, two-sided unpaired t -test. **b**, mRNA expression analysis in 20-week-old L-dKO tumours ($n=3$) compared to livers from age-matched control mice ($n=3$). *CD90* (also known as *Thy1*; * $P=0.0104$), *CD133* (also known as *Prom1*; * $P=0.0227$), *Cd44* (* $P=0.0473$) and *Aldh1* (also known as *Aldh1a1*; * $P=0.0107$), two-sided unpaired t -test. Data are mean \pm s.d. **c**, Diagram showing the scheme used for mass spectrometry-based proteomic analysis. Hepatic proteomes from control mice ($n=6$) were

combined and used as a pooled control against tumours ($n=12$) from four different L-dKO mice. **d**, Graph showing number of proteins upregulated (red), downregulated (blue) and not regulated (grey) in 12 tumours from **c**. ANOVA-based two-sample t -test with a FDR of 2%, $S_0=1$ was used to determine regulation ($n=12$). **e**, List of kinases and phosphatases (including putative phosphatases) regulated in a minimum of 10 out of 12 tumours. **f**, *Lhpp* mRNA expression analysis in 20-week-old L-dKO tumours compared to livers from age-matched control mice ($n=4$). $P=0.0022$, two-sided unpaired t -test. Data are mean \pm s.d.



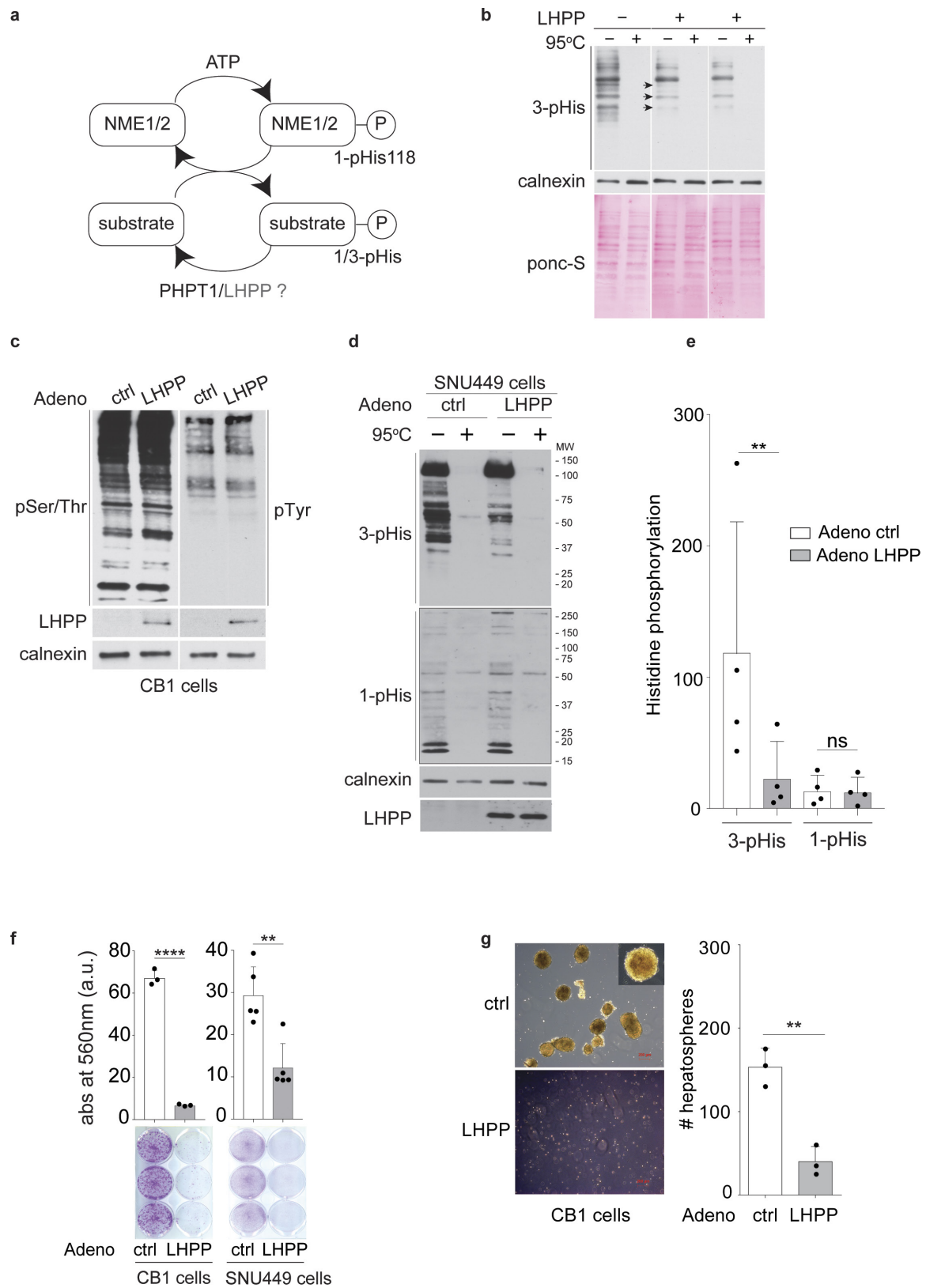
Extended Data Figure 2 | LHPP protein is conserved in eukaryotes. a, LHPP protein is highly conserved in eukaryotes, from worm to human. **b**, Human LHPP protein shows weak homology to known human histidine phosphatases PHPT1 and PGAM5.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Loss of LHPP expression is confined to L-dKO tumours. **a**, Quantification of immunoblot (from Fig. 2a) indicates reduced LHPP ($***P=0.0004$) and increased NME1 ($*P=0.0254$) and NME2 ($*P=0.0105$) expression in tumours compared to age-matched control littermates (band intensities in each lane are normalized to intensity of corresponding calnexin protein levels), control mice ($n=4$), L-dKO tumours ($n=8$). P values are from a two-sided unpaired t -test. Data are mean \pm s.d. **b**, Immunoblot analysis of liver tumour ($n=6$) and non-tumour ($n=3$) tissue from three independent 20-week-old L-dKO mice shows tumour-specific loss of LHPP expression. Ponceau S image of the whole blot is also included. For full scan, see Supplementary Fig. 3. **c**, Immunohistochemistry analyses showing increased NME1

expression in L-dKO liver compared to liver from control mice. Although L-dKO tumours displayed reduced expression of LHPP, hepatic LHPP expression in the non-tumour region was comparable to control mice. Parallel tissue sections that were processed similarly (without primary antibody) served as antibody control. Similar results were obtained in livers from 4 independent control mice and L-dKO mice. **d**, Immunoblot analysis of liver from control mice and L-dKO liver tumours treated with INK128 (1 mg kg^{-1} body weight; two intraperitoneal injections in 24 h; $n=4$ mice per condition, lysates from two mice are pooled into one lane). LHPP expression remains unchanged, but NME1 and NME2 expression is reduced considerably in L-dKO mice after INK128 treatment. For full scans, see Supplementary Fig. 4.

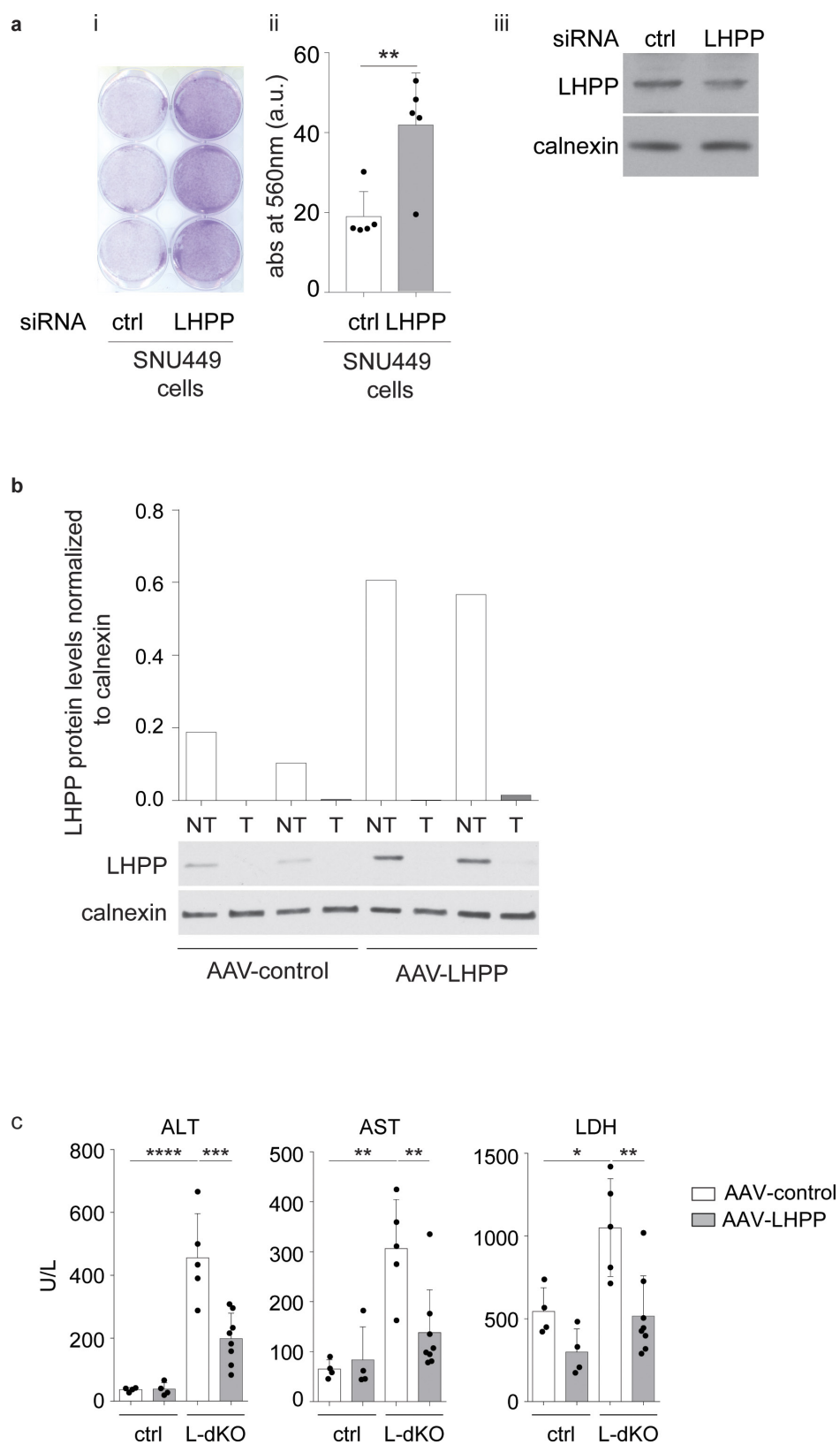


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | LHPP is a protein histidine phosphatase.

a, Diagram of NME1 and NME2 autophosphorylation on His118. The high-energy phosphate is later transferred from the histidine kinase dimer pair to the histidine on the substrate. **b**, CB1 cell lysate was electrophoresed and transferred onto a nitrocellulose membrane. Two sets of the membrane were treated with recombinant full-length LHPP (for 4 h) in TMD buffer (see Methods) and the other set of the membrane was treated with plain TMD buffer (pH 8.5). Immunoblotting analysis of both the membranes with 3-pHis antibody (clone SC44-1) indicates that LHPP is a protein pHis phosphatase. For full scans, see Supplementary Fig. 6. **c**, Immunoblotting shows no apparent change pSer/Thr or pTyr levels in CB1 cells infected with adenovirus containing LHPP compared to cells infected with RFP-only control. Similar result was observed in two independent experiments. For full scans, see Supplementary Fig. 9. **d**, **e**, Immunoblotting (**d**) and quantification (**e**) show a reduction in 3-pHis levels in SNU449 cells infected with adenovirus containing LHPP compared to control-infected cells. Similar results were observed in four independent biological experiments.

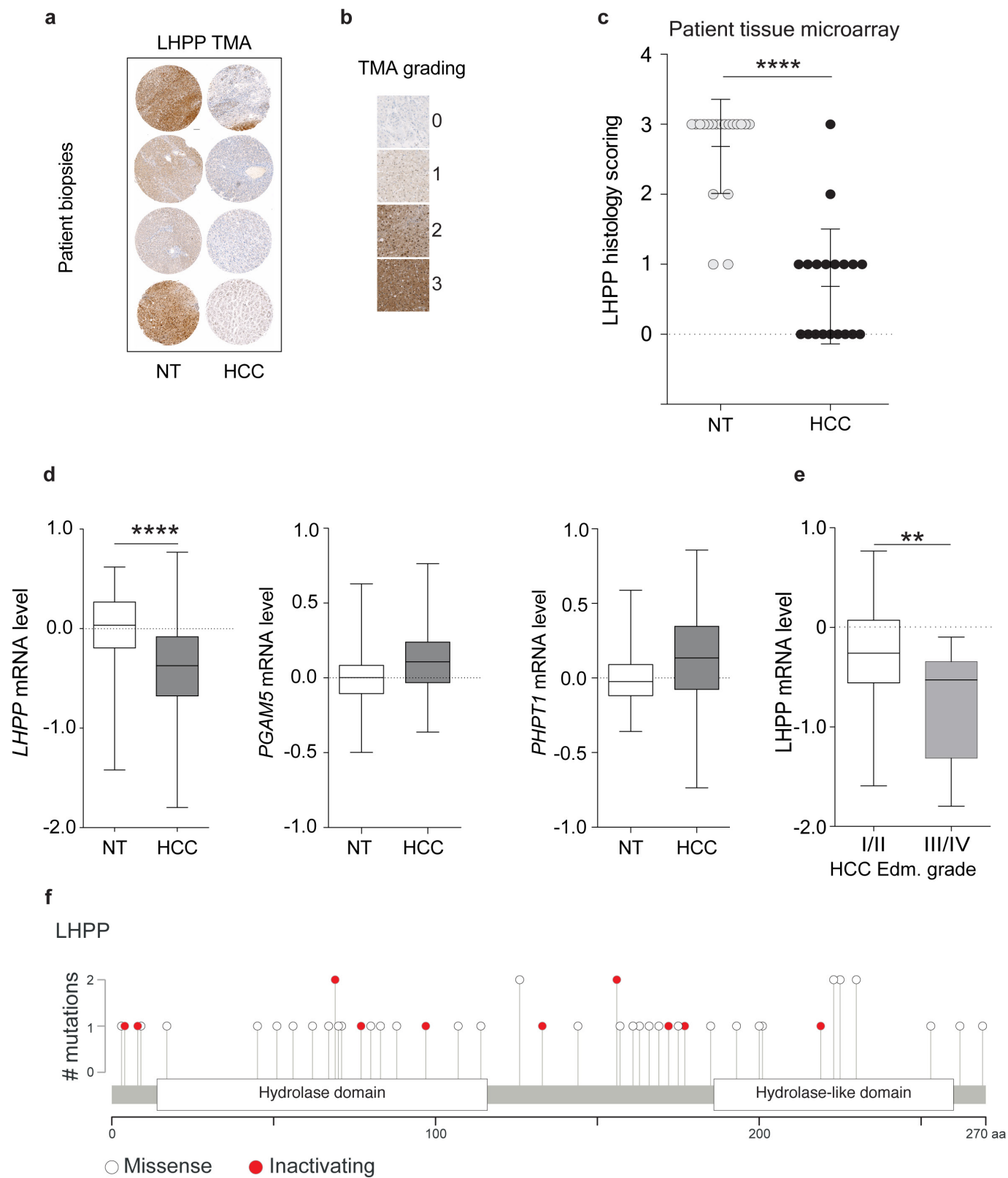
For full scans, see Supplementary Fig. 8. A 50–60% confluent dish was first infected, 24 h after infection, the cells were trypsinized and 2×10^5 cells were seeded into a 6-well adherent tissue culture dish and 24 h later lysed in pHis compatible buffer (see Methods). The monoclonal antibodies used to detect 3-pHis and 1-pHis were SC44-1 and SC1-1, respectively ($n = 4$). $^{***}P = 0.0019$, two-sided ratio paired t -test. Data are mean \pm s.d. **f**, Colony-forming assay in CB1 cells ($n = 3$ biologically independent experiments) and SNU449 ($n = 5$ biologically independent experiments) cells shows reduced proliferation upon LHPP over expression. Twelve hours after infection, CB1 (2×10^4 per well) and SNU449 (4×10^4 per well) cells were seeded into a 6-well plate tissue culture plate and stained with crystal violet (2% crystal violet in 20% methanol) on day 7. $^{**}P = 0.0026$, $^{****}P < 0.0001$, two-sided unpaired t -test. Data are mean \pm s.d. See Methods for further details. **g**, Photomicrographs of LHPP-overexpressing CB1 cells (left) displayed reduced ability to form hepatospheres in ultralow attachment plates; $n = 3$ biologically independent experiments. $^{**}P = 0.0024$, two-sided unpaired t -test. Data are mean \pm s.d. (see Methods).



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | LHPP is a tumour suppressor. **a**, Left and middle, colony-forming assay in SNU449 cells shows increased proliferation upon LHPP knockdown using pooled short interfering RNA (siRNA). Then 36 h after infection, SNU449 cells were re-transfected with LHPP siRNA and 4×10^4 cells per well were seeded into a 6-well tissue culture plate and allowed to proliferate for another 3 days. At the end of the third day, cells were stained with crystal violet (2% crystal violet in 20% methanol), and similar results were obtained from five biologically independent experiments. $**P = 0.0075$, two-sided unpaired *t*-test. Data are mean \pm s.d. Right, immunoblot analysis ($n = 1$) revealed a reduction in LHPP protein levels after siRNA transfection. Longer exposure of the LHPP blot relative to the blot in Extended Data Fig. 4d. In addition to immunoblotting, a reduction in *LHPP* transcript levels after siRNA transfection was also confirmed in two independent experiments (data not shown). For full scans, see Supplementary Fig. 10. **b**, Immunoblot

analysis (bottom) and quantification (top) of non-tumour and tumour liver lysates confirmed LHPP overexpression in 20-week-old AAV-LHPP-infected mice (as in Fig. 3e). For full scans, see Supplementary Fig. 11. **c**, Serum extracted from mice in (Fig. 3e) and analysed for ALT, AST and LDH levels shows a reduction in liver damage upon AAV-LHPP infection, unit per litre (U l^{-1}) (control mice infected with AAV-control and AAV-LHPP virus ($n = 4$), L-dKO mice infected with AAV-control ($n = 5$) and AAV-LHPP virus ($n = 8$)). Data are mean \pm s.d. *P* values are by one-way ANOVA with Tukey's multiple-comparison test. ALT (adjusted $***P = 0.004$, $****P < 0.0001$), AST (adjusted $**P = 0.013$ for control mice versus L-dKO mice transfected with AAV-control; adjusted $**P = 0.0075$ for L-dKO mice transfected with AAV-control versus L-dKO mice transfected with AAV-LHPP), LDH (adjusted $*P = 0.0196$, $**P = 0.0037$).



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | LHPP is a tumour suppressor in human HCC.

a, Representative immunohistochemistry images from the LHPP-stained tissue microarray. Scale bar, 100 μ m. **b**, Representative images and staining intensities of LHPP staining in the tissue microarray. **c**, Tissue microarray (TMA) indicates that LHPP is significantly downregulated in human HCC compared to adjacent non-tumour tissue ($n = 20$ HCC patients, **** $P < 0.0001$, two-sided paired t -test). **d**, Box plots showing a significant reduction in LHPP expression in HCC tissue compared to adjacent non-tumour tissue ($n = 59$). **** $P < 0.0001$, two-sided Mann–Whitney U test. Non-tumour (minimum, -1.422 , median, 0.03329 , maximum, 0.6179 , lower 95% confidence interval (CI) of mean -0.0931 , upper 95% CI of mean 0.0931) and HCC (minimum, -1.797 , median, -0.3729 , maximum, 0.7683 , lower 95% CI of mean -0.5328 , upper 95% CI of mean -0.2506). Expression of *PHPT1* and *PGAM5* was not reduced

in HCC. **e**, Box plots showing a strong reduction in LHPP expression in Edmondson grade III/IV tumours (aggressive tumours, $n = 17$) compared to Edmondson grade I/II tumours (less aggressive tumours, $n = 42$). ** $P = 0.0036$, two-sided Mann–Whitney U test. Edmondson grade I/II (minimum, -1.592 , median, -0.2594 , maximum 0.7683 , lower 95% CI of mean -0.3987 , upper 95% CI of mean -0.0986) and III/IV (minimum, -1.797 , median, -0.5281 , maximum, -0.09671 , lower 95% CI of mean -1.018 , upper 95% CI of mean -0.4717). **f**, Lollipop plot showing the distribution of mutations in LHPP across pan-cancer datasets in the TCGA and ICGC portal. Solid white circles indicate missense mutations, red circles indicate inactivating mutations (including nonsense mutations, splice site mutations and frameshift deletions; detailed list of mutations is provided in Supplementary Table 5).

Optogenetic regulation of engineered cellular metabolism for microbial chemical production

Evan M. Zhao¹, Yanfei Zhang¹, Justin Mehrl¹, Helen Park¹, Makoto A. Lalwani¹, Jared E. Toettcher² & José L. Avalos^{1,3}

The optimization of engineered metabolic pathways requires careful control over the levels and timing of metabolic enzyme expression^{1–4}. Optogenetic tools are ideal for achieving such precise control, as light can be applied and removed instantly without complex media changes. Here we show that light-controlled transcription can be used to enhance the biosynthesis of valuable products in engineered *Saccharomyces cerevisiae*. We introduce new optogenetic circuits to shift cells from a light-induced growth phase to a darkness-induced production phase, which allows us to control fermentation with only light. Furthermore, optogenetic control of engineered pathways enables a new mode of bioreactor operation using periodic light pulses to tune enzyme expression during the production phase of fermentation to increase yields. Using these advances, we control the mitochondrial isobutanol pathway to produce up to $8.49 \pm 0.31 \text{ g l}^{-1}$ of isobutanol and $2.38 \pm 0.06 \text{ g l}^{-1}$ of 2-methyl-1-butanol micro-aerobically from glucose. These results make a compelling case for the application of optogenetics to metabolic engineering for the production of valuable products.

Metabolic engineering aims to rewire the metabolism of organisms for efficient conversion of inexpensive substrates into valuable products such as chemicals, fuels, or drugs^{1,2}. Fine-tuning the timing and levels of expression of enzymes involved in both natural and engineered pathways can relieve bottlenecks and minimize the metabolic burden of chemical production^{3,4}. This is especially critical when the product of interest or its precursors are toxic, or when the biosynthetic pathway of interest competes with endogenous pathways that are essential for cell growth.

To address these challenges, metabolic engineers frequently use inducible systems to control metabolic enzyme expression^{5–7} (see Supplementary Discussion). This approach makes it possible to separate bioreactor operation into two phases: a growth phase, during which product biosynthesis is repressed, and a production phase, when flux through the engineered pathway is maximized. Essential pathways that compete with product formation can be controlled with ‘metabolic valves’—genetic programs that express essential enzymes during the growth phase to build biomass, and repress them during the production phase to redirect metabolism towards desired products^{5,8}.

Light is an attractive strategy to control gene expression in yeast for metabolic engineering applications. It is inexpensive and compatible with any carbon source or nutrient composition. Furthermore, light can be applied or removed instantaneously; this precise control over the level or duration of enzyme expression could simplify the screening of optimal proportions of metabolic pathway enzymes and enable new time-varying modes of control^{9,10}. Light-switchable transcription modules have been shown to enable non-toxic, tunable gene expression in a variety of organisms^{11,12}, including yeast^{11,13–15}. We thus sought to test whether optogenetics could be used to control rewired cellular metabolisms to overproduce valuable chemicals.

Here, we describe two powerful optogenetic gene expression systems for yeast, OptoEXP and OptoINVRT, based on the blue light-activated EL222 gene expression system¹². Using these systems, we show that it is possible to activate and to repress distinct sets of genes in a light-dependent manner. We apply our approach to control endogenous and engineered metabolic pathways to define the growth and production phases of fermentation, enabling production of three valuable chemicals (lactate, isobutanol and 2-methyl-1-butanol (2-MBOH)), the biosyntheses of which directly compete with essential ethanol production. Using a time-varying illumination schedule, we achieve titres of $8.49 \pm 0.31 \text{ g l}^{-1}$ (mean \pm s.d.) of isobutanol and $2.38 \pm 0.06 \text{ g l}^{-1}$ of 2-MBOH. Our results thus reveal that a simple technology—bidirectional light-controlled enzyme expression—offers a rich set of tools for metabolic engineering.

Our first goal was to construct bidirectional gene circuits in yeast to either induce or repress genes of interest with light. We used the EL222 optogenetic transcription system, which consists of a light-sensitive transcription factor from *Erythrobacter litoralis* (EL222) and its corresponding C120 promoter (P_{C120}), to drive the expression of a gene of interest^{16–18} (Fig. 1a). The EL222 system is a robust and versatile gene expression platform, previously applied in *Escherichia coli*, mammalian cells and zebrafish^{12,19,20}. We first constructed a yeast strain, YEZ139, in which expression of VP16–EL222 (a fusion of EL222 with the transcriptional activation domain of VP16 and a nuclear localization signal) was driven by the strong constitutive promoter of *PGK1* (P_{PGK1}) (P_{PGK1} –VP16–EL222) and expression of green fluorescent protein (GFP) was driven by the P_{C120} promoter (P_{C120} –GFP; see Methods, Supplementary Tables 1, 2, Extended Data Fig. 1). We called this system (and variants with different promoters driving EL222) OptoEXP.

OptoEXP enables strong and titratable light-inducible gene expression. In both glucose and glycerol media, cells with OptoEXP controlling GFP expression show a 43-fold increase in GFP expression when exposed to constant light compared to cells incubated in the dark, whereas intermittent light pulses produce intermediate expression levels (Fig. 1b, Extended Data Fig. 2). We found that all light sources used were sufficiently bright to activate the EL222 system maximally; thus, varying duty cycle was a reliable and reproducible method for achieving intermediate gene expression output (Extended Data Fig. 3, Methods). The maximum activation levels of OptoEXP are comparable to those reached by the *ADH1* promoter (P_{ADH1}), a constitutive promoter commonly used in metabolic engineering (Fig. 1c).

To construct a light-repressible gene circuit, we inverted the response of the OptoEXP system in a manner akin to the NOT logic gate used in digital processes. We harnessed the yeast galactose (GAL) regulon²¹, in which Gal80 binds to and inhibits the Gal4 transcription factor, blocking its ability to induce expression from the *GAL1* promoter (P_{GAL1}). We reasoned that engineering yeast cells with constitutive expression of *GAL4* and *GAL80* expression under the control of VP16–EL222 would lead to constitutive expression from

¹Department of Chemical and Biological Engineering, Hoyt Laboratory, Princeton University, 25 William Street, Princeton, New Jersey 08544, USA. ²Department of Molecular Biology, 140 Lewis Thomas Laboratory, Washington Road, Princeton, New Jersey 08544, USA. ³The Andlinger Center for Energy and the Environment, Princeton University, 86 Olden Street, Princeton, New Jersey 08544, USA.

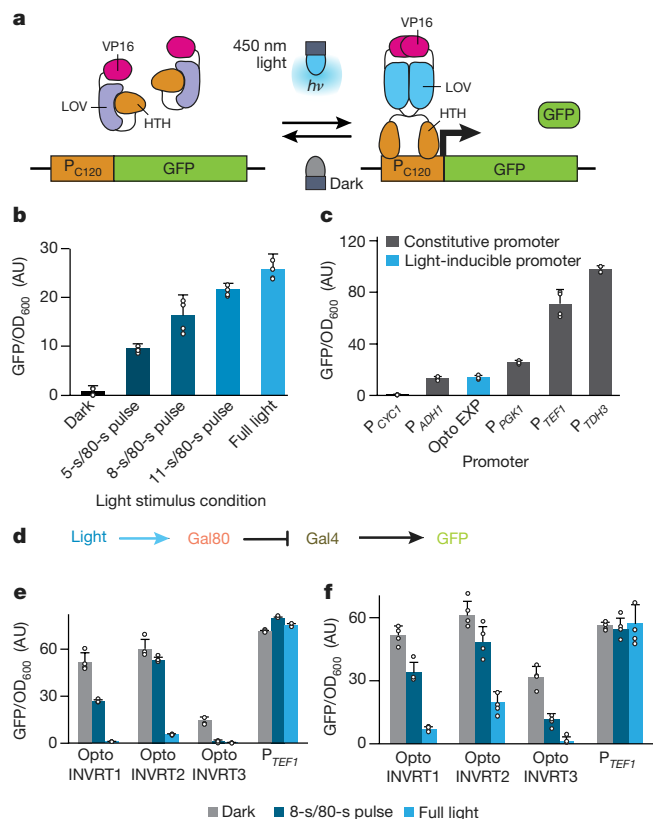


Figure 1 | Characterization of optogenetic circuits. **a**, Reversible OptoEXP system based on VP16-EL222 that is sensitive to 450 nm light. $h\nu$ indicates the energy of photons. THH, helix–turn–helix DNA-binding domain. **b**, GFP expression with OptoEXP (YEZ139) tuned with different duty cycles, for example 5-s/80-s pulse means that the light was on for 5 s and off for 75 s. AU, arbitrary units of fluorescence and optical density. **c**, Maximum GFP expression from OptoEXP compared to common constitutive promoters using strains YEZ27 (*CYC1* promoter, *P_{CYC1}*), YEZ28 (*P_{ADH1}*), YEZ29 (*P_{PGK1}*), YEZ30 (*P_{TDH3}*), YEZ31 (*P_{TEF1}*) and YEZ32 (*P_{C120}*). **d**, OptoINVRT circuit diagram (details in Extended Data Fig. 4a, b). **e**, OptoINVRT circuits in CENPK.2-1C-derived (*GAL80Δ*, *GAL4Δ*) strains YEZ100 (OptoINVRT1), YEZ101 (OptoINVRT2) and YEZ102 (OptoINVRT3). **f**, OptoINVRT circuits in BY4741-derived (*GAL80Δ*, *PDC1Δ*, *PDC5Δ*, *PDC6Δ*) strains YEZ115 (OptoINVRT1), YEZ116 (OptoINVRT2) and YEZ117 (OptoINVRT3). All data are shown as mean values; dots represent individual data points; error bars represent the s.d. of four biologically independent 1-ml sample replicates exposed to the same conditions. All experiments were repeated at least three times.

the *P_{GAL1}* promoter in the dark and repression of *P_{GAL1}* in the light (Fig. 1d, Extended Data Fig. 4a, b).

Starting from a strain in which both *GAL80* and *GAL4* are deleted, YEZ44, we constructed three variants of this core inverter topology, which we termed OptoINVRT1, OptoINVRT2 and OptoINVRT3 (Supplementary Tables 1, 2). These variants differed in the strength of the promoter driving *GAL4* and the fusion of a photosensitive degron (PSD) domain²² to the C terminus of Gal4 to induce faster and more complete light-dependent repression (Extended Data Fig. 4a, b and Supplementary Table 3). Using *P_{GAL1}*–GFP as a reporter, we found that all three OptoINVRT circuits exhibit robust light-induced gene repression (Fig. 1e, Supplementary Table 4). OptoINVRT2 has the highest maximum expression in the dark (in YEZ101, almost 85% of *TEF1* promoter (*P_{TEF1}*) levels), while OptoINVRT3 has the highest levels of repression (in YEZ102, more than 70-fold) and lowest expression in full light. All three OptoINVRT circuits show similar responses in a second yeast strain, Y202, which is relevant for metabolic engineering owing to the deletion of all three of its pyruvate decarboxylase genes (strain S288C with the

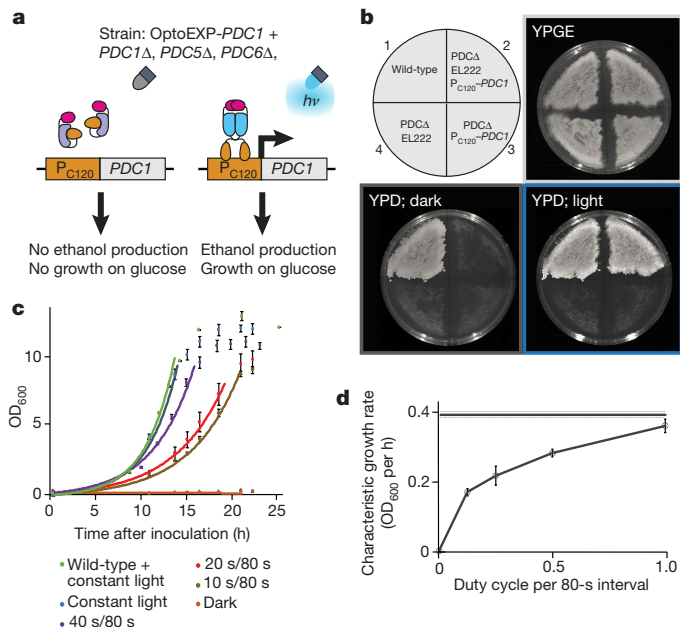


Figure 2 | A light-dependent metabolic valve for ethanol production.

a, Using OptoEXP to build light-inducible metabolic valve for *PDC1*. **b**, Light-dependent growth on glucose of a triple PDC mutant strain containing the OptoEXP-*PDC1* metabolic valve. Top left, wild-type BY4741; top right, YEZ61-23 (*PDC1Δ*, *PDC5Δ*, *PDC6Δ*, *P_{TEF1}*-VP16-EL222, *P_{C120}*-*PDC1*); bottom left, YEZ50C (*PDC1Δ*, *PDC5Δ*, *PDC6Δ*, *P_{TEF1}*-VP16-EL222); bottom right, YEZ61C (*PDC1Δ*, *PDC5Δ*, *PDC6Δ*, *P_{C120}*-*PDC1*). **c**, YEZ61-23 and BY4741 growth in liquid synthetic complete medium supplemented with 2% glucose at different light doses, with best fits to exponential growth curves. **d**, Maximum growth rate of YEZ61-23 as a function of light. Solid line represents the mean wild-type (BY4741) growth rate (dotted lines indicate the s.d.). All data are shown as mean values (values for individual points are available in the Source Data for this figure); error bars represent the s.d. of three biological replicates exposed to the same conditions. All experiments were repeated at least three times.

following mutations: *GAL80Δ*, *PDC1Δ*, *PDC5Δ* and *PDC6Δ*) (Fig. 1f, Supplementary Table 4). Our results thus demonstrate that the OptoINVRT platform can achieve a wide range of expression levels, light sensitivities and fold-change responses in different strain backgrounds for flexible incorporation in diverse metabolic engineering applications.

To reduce the ethanol by-product in pyruvate-derived chemical biosyntheses, while still allowing cell growth on glucose, we used OptoEXP to control pyruvate decarboxylation (PDC), an essential step in ethanol biosynthesis. Completely removing PDC activity in *S. cerevisiae* (by deleting *PDC1*, *PDC5* and *PDC6*) does not suffice, because the triple deletion renders yeast unable to grow on glucose owing to the essential role of these genes in NAD⁺ recycling for glycolysis²³ and to the lack of effective alternatives for ATP generation owing to glucose-mediated repression of respiration²⁴. Thus, we used our OptoEXP circuit to build a light-dependent metabolic valve for *PDC1* expression (Fig. 2a). Light stimulation ‘opens’ the valve, enabling robust cell growth and ethanol production; dark incubation ‘closes’ the valve to limit the metabolism of pyruvate into ethanol, thereby inhibiting cell growth on glucose, and enhancing the biosynthesis of alternative pyruvate-derived products.

A strain with all PDC genes deleted and with OptoEXP driving *PDC1* expression, YEZ61-23 (Supplementary Table 2), exhibits light-dependent growth on glucose. YEZ61-23 grows on glycerol with ethanol (yeast peptone with 3% glycerol and 2% ethanol (YPGE)) independently of light because PDC is dispensable for respiration. However, it is only able to grow on glucose plates (yeast peptone with 2% glucose (YPD)) in the presence of blue light (Fig. 2b), with

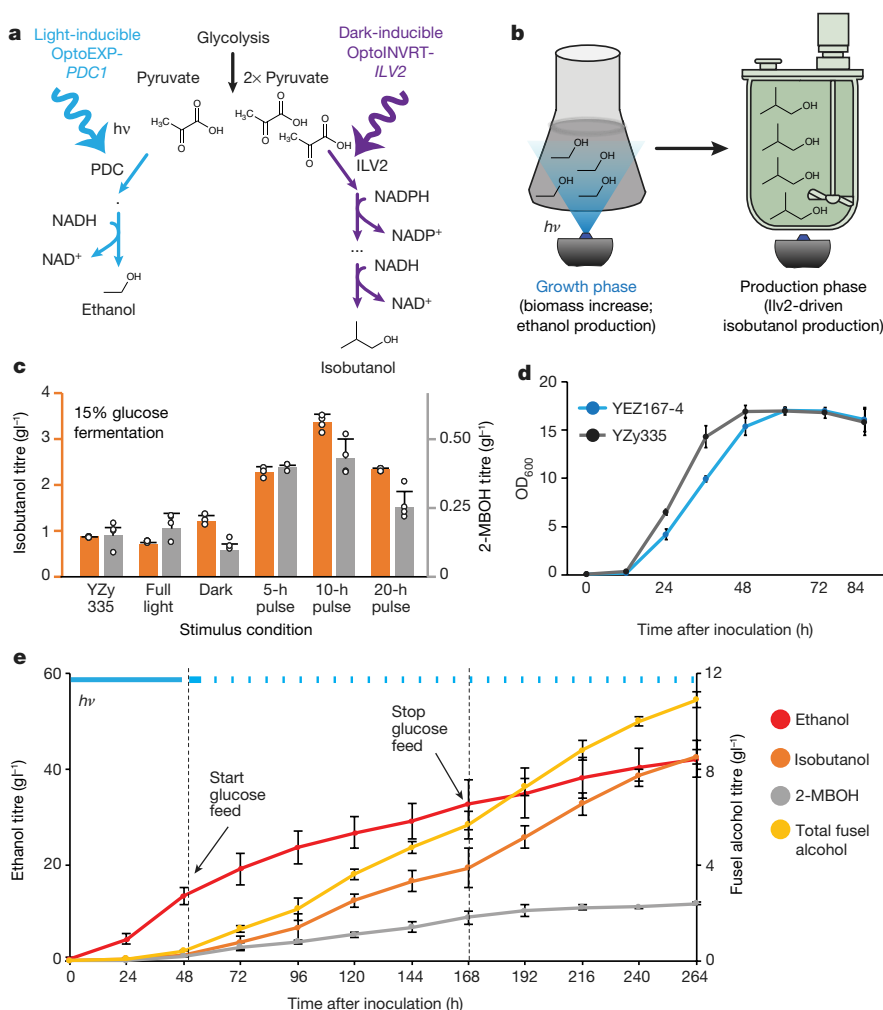


Figure 3 | Light-controlled isobutanol production. **a**, Ethanol and isobutanol biosynthesis controlled by OptoINVIRT-*ILV2* and OptoEXP-*PDC1*, respectively. **b**, Blue light allows growth of YEZ167-4 in glucose, as well as ethanol formation, whereas dark conditions initiate the production phase by stopping *PDC1* induction and de-repressing *ILV2*. **c**, Isobutanol and 2-MBOH production of YEZ167-4 in fermentations of 80 h in 150 g l⁻¹ glucose and blue light pulses of 15 s on and 65 s off for 30 min every 5, 10, or 20 h. YZy335 is a light-insensitive isobutanol-producing control strain. **d**, Growth of YEZ167-4 in a 2-l bioreactor and under constant blue light, compared to the YZy335 light-insensitive control. **e**, Isobutanol and

2-MBOH production of YEZ167-4 in a 0.5-l fed-batch pH-controlled fermenter using periodic light stimulation during the production phase. Total fusel alcohols represent the sum of isobutanol and 2-MBOH. Blue bar indicates the time of blue light exposure. Dotted lines indicate the start and stop of the glucose feed. All data are shown as mean values; dots in **c** represent individual data points; individual points for **d** and **e** are available in the Source Data for this figure and Supplementary Table 5, respectively; error bars represent the s.d. of three biological replicates (**c**) or three separate bioreactor runs (**d**, **e**). All experiments were repeated at least twice.

its growth rate in liquid medium reaching $91.9 \pm 4.9\%$ of the wild type (BY4741) at full light (Fig. 2c, d). Substantial growth is achieved even at low light doses ($43.7 \pm 2.7\%$ of wild-type growth rate with only 12.5% light duty cycle), suggesting that light-dependent growth may be achieved at cell densities relevant for metabolic engineering applications.

We then tested whether the bidirectional control afforded by combining our OptoEXP and OptoINVIRT circuits could be used to drive two phases of cellular metabolism: a growth phase with mainly ethanol fermentation and a production phase in which carbon accumulates as a desired valuable product instead (either lactate or isobutanol). Production of lactate or isobutanol competes directly with ethanol production as distinct enzymes modify pyruvate to produce each product (pyruvate decarboxylase 1 (Pdc1) in the case of ethanol, lactate dehydrogenase (Ldh) in the case of lactate, and acetolactate synthase (Ilv2) and subsequent enzymes in the case of isobutanol) (Fig. 3a, Extended Data Fig. 5a). We reasoned that by controlling the expression of *PDC1* with OptoEXP and the expression of Ldh or *ILV2* with OptoINVIRT, moving cells from light to dark would shift

the metabolism from the production of ethanol to the production of lactate or isobutanol (Fig. 3b, Extended Data Fig. 5b).

We first tested our strategy in the simpler case of lactate biosynthesis, which requires overexpression of only one enzyme (Ldh). The presence of lactate is also well tolerated by yeast. We integrated multiple copies of *P_{GAL1}-Ldh* and *P_{CI20}-PDC1* into our Y202-derived OptoINVIRT strains (Extended Data Fig. 6a) and observed a pronounced increase in lactate biosynthesis upon a shift from dark to light, consistent with light-dependent changes in *PDC1* and Ldh expression levels (Methods, Extended Data Figs 4c, 5). On the basis of these results, we next tested our optogenetic circuits on the more complex biosynthesis of isobutanol (Fig. 3a, b), an advanced biofuel with much higher toxicity than lactate. Our approach was to optogenetically control only the first enzyme in the isobutanol pathway (acetolactate synthase), reasoning that if subsequent enzymes were constitutively expressed, the overall pathway flux would still be light-controlled.

We transformed Y202 with OptoEXP, each one of the OptoINVIRT circuits, multiple copies of *P_{GAL1}-ILV2* and *P_{CI20}-PDC1* and the rest of the mitochondrial isobutanol biosynthetic pathway under strong

constitutive promoters²⁵, to produce strains YEZ159, YEZ156 and HPY6 (Supplementary Tables 1, 2, Methods). We found that colonies from YEZ159, containing OptoINVERT1, produced the highest isobutanol titres from 4% glucose (Extended Data Fig. 4d, e). To enhance isobutanol production further, we deleted the mitochondrial branched chain amino acid aminotransferase (encoded by *BAT1*, also known as *YBT1*), which competes for the α -ketoisovalerate precursor^{26,27}, resulting in strain YEZ167-4 (Supplementary Table 2). YEZ167-4 contains six copies of *P_{C120}-PDC1* (Extended Data Fig. 6a), leading to fast cell growth under full light.

By varying the cell density at which cultures are switched from dark to light and the incubation time in the dark before starting the fermentation (Extended Data Fig. 5c), we found that YEZ167-4 can produce as much as $735 \pm 15 \text{ mg l}^{-1}$ of isobutanol from 2% glucose over 48 h (a yield of $34.2 \pm 0.7 \text{ mg isobutanol g}^{-1} \text{ glucose}$, Extended Data Fig. 7). Switching to 15% glucose and fermentation periods of 80 h only increases production to $1.22 \pm 0.11 \text{ g l}^{-1}$ of isobutanol (Fig. 3c). However, under these conditions YEZ167-4 is unable to consume all the glucose in the medium (Extended Data Fig. 8a), indicating a stalled fermentation in which cells have undergone premature metabolic arrest.

We hypothesized that during prolonged dark incubation, *Pdc1* becomes limiting to a point at which cellular metabolism arrests owing to NAD^+ depletion. We reasoned that periodic pulses of light during the production phase of fermentation could transiently induce *PDC1* expression, thus increasing NAD^+ pools and restoring cellular metabolism, glucose consumption and isobutanol production. We tested different light schedules, applying periodic illumination bouts (30 min of blue light at a duty cycle of 15 s on and 65 s off once every 5, 10 or 20 h throughout the production phase of the fermentation, which lasted 80 h). When cells are exposed to one light bout every 10 h, isobutanol production nearly triples to $3.37 \pm 0.17 \text{ g l}^{-1}$ (Fig. 3c), and 2-MBOH, another desirable advanced biofuel co-produced with the mitochondrial isobutanol pathway²⁵, is produced at $433 \pm 69 \text{ mg l}^{-1}$. These isobutanol yields are more than 2.5 times higher than if cells are kept in the dark throughout the production phase of the fermentation, and four times higher than YZy335, a strain that contains the entire isobutanol pathway and *BAT1* deletion but lacks light-inducible metabolic control. We directly confirmed that our light pulses partially replenish the NAD^+ metabolite pool under these fermentation conditions (Methods, Extended Data Fig. 8b).

To determine the applicability of optogenetics to metabolic engineering in laboratory-scale bioreactors, we measured the growth of the OptoEXP-driven *PDC1* strain, YEZ167-4, in a 2-l fermenter with 15% glucose and constant blue light (Extended Data Fig. 9a, b). Under these conditions, YEZ167-4 is able to reach the same optical density at 600 nm ($\text{OD}_{600} = 17.1$) as a light-independent, *PDC*⁺ strain, YZy335 (Supplementary Table 2), indicating that light penetration into the bioreactor is sufficient to drive robust *PDC1* expression and growth (Fig. 3d). We then measured the production of isobutanol and 2-MBOH in fed-batch 0.5-l bioreactors under microaerobic conditions and pH control (Extended Data Fig. 9c, d). We used constant illumination to build YEZ167-4 biomass during a batch growth phase, and then switched to a fed-batch production phase with periodic bouts of light to maintain sufficient NAD^+ levels to preserve metabolic activity. This fermentation strategy yielded $8.49 \pm 0.31 \text{ g l}^{-1}$ of isobutanol and $2.38 \pm 0.06 \text{ g l}^{-1}$ of 2-MBOH, with post-induction average yields of $53.5 \pm 8.4 \text{ mg g}^{-1}$ glucose for isobutanol (13% of theoretical maximum) and $14.17 \pm 2.57 \text{ mg g}^{-1}$ glucose of 2-MBOH, with measured instantaneous isobutanol yields of up to 270.6 mg g^{-1} glucose (66% of theoretical yield) (Fig. 3e, Supplementary Table 5, Methods, Supplementary Results). Our mechanistic predictions of a darkness-induced metabolic shift to isobutanol biosynthesis are supported by quantitative PCR (qPCR) measurements of *PDC1* and *ILV2* mRNA levels (Extended Data Fig. 6b).

Optogenetics is a powerful way to control cellular physiology. Photosensitive proteins have had an enormous impact in neuroscience,

cell biology and developmental biology by controlling ion channels, enzyme activity and gene expression^{9,28–30}. Our work demonstrates that optogenetics also holds promise in metabolic engineering, enabling the reversible control and fine-tuning of engineered metabolic pathways.

Nevertheless, combining optogenetics and metabolic engineering has its challenges. The high cell densities usually associated with microbial fermentations might be predicted to severely limit light penetration, which is one reason we initially adopted our design of a darkness-induced production phase. However, additional experiments showed robust light-stimulated gene expression could be achieved even at cell densities as high as $\text{OD}_{600} = 50$ in 5-l bioreactors (Extended Data Fig. 10), suggesting that the high light sensitivity and strong levels of gene expression achieved by the current generation of optogenetic tools may be sufficient even in these challenging conditions. Even more potent responses to dim light could potentially be achieved by using well-characterized light-oxygen-voltage-sensing-domain (LOV) mutants with longer lit-state lifetimes¹⁸ (our current VP16-EL222 remains in the photoactivated state for approximately 30 s; ref. 12), or other photoactive proteins with highly stable light-switchable conformations¹³ (see Supplementary Discussion). For fermentation conditions that pose a greater challenge to light penetration than in this study, our OptoINVERT circuits offer a solution to control gene expression as they do not require light during the fermentation.

We used optogenetic regulation of engineered metabolic pathways to address the long-standing challenge of ethanol competition in branched-chain alcohol production. We and others have developed triple *PDC* deletion strains (*PDC1* Δ , *PDC5* Δ , *PDC6* Δ) that recovered their ability to grow on glucose by directed evolution²³. However, efforts to produce isobutanol in these strains from mitochondrial (unpublished results) or cytosolic³¹ biosynthetic pathways have thus far been unsuccessful. Our light-controlled metabolic valve offers an effective alternative to genetically deleting essential pathways that compete with a pathway of interest⁵. This allowed us to surpass the highest titres of isobutanol and 2-MBOH reported for yeast in the peer-reviewed literature that we have analysed by fivefold (Supplementary Table 6) and 20-fold²⁵, respectively, making a strong case for the application of optogenetics to metabolic engineering.

Optogenetic regulation of metabolic pathways enables new strategies for optimizing engineered pathways and fermentation conditions using periodic light pulses. Although further development for industrial applications is needed, in the future, light inputs delivered to a bioreactor could be automatically controlled in response to feedback from different fermentation outputs (for example, optical density or the readout of biosensors), providing unprecedented capabilities for operating, optimizing and automating fermentations for valuable product biosynthesis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 May 2017; accepted 13 February 2018.

Published online 21 March 2018.

1. Woolston, B. M., Edgar, S. & Stephanopoulos, G. Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* **4**, 259–288 (2013).
2. Zhang, Y. *et al.* Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and Mammalian cells. *J. Am. Chem. Soc.* **128**, 13030–13031 (2006).
3. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355–1358 (2010).
4. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).
5. Tan, S. Z., Manchester, S. & Prather, K. L. J. Controlling central carbon metabolism for improved pathway yields in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **5**, 116–124 (2015).
6. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
7. Gu, P., Su, T., Wang, Q., Liang, Q. & Qi, Q. Tunable switch mediated shikimate biosynthesis in an engineered non-auxotrophic *Escherichia coli*. *Sci. Rep.* **6**, 29745 (2016).

8. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng.* **28**, 104–113 (2015).
9. Toettcher, J. E., Voigt, C. A., Weiner, O. D. & Lim, W. A. The promise of optogenetics in cell biology: interrogating molecular circuits in space and time. *Nat. Methods* **8**, 35–38 (2011).
10. Miliadis-Argeitis, A. *et al.* *In silico* feedback for *in vivo* regulation of a gene expression circuit. *Nat. Biotechnol.* **29**, 1114–1116 (2011).
11. Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nat. Methods* **7**, 973–975 (2010).
12. Motta-Mena, L. B. *et al.* An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat. Chem. Biol.* **10**, 196–202 (2014).
13. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nat. Biotechnol.* **20**, 1041–1044 (2002).
14. Taslimi, A. *et al.* Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nat. Chem. Biol.* **12**, 425–430 (2016).
15. Salinas, F., Rojas, V., Delgado, V., Agosin, E. & Larrondo, L. F. Optogenetic switches for light-controlled gene expression in yeast. *Appl. Microbiol. Biotechnol.* **101**, 2629–2640 (2017).
16. Nash, A. I. *et al.* Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc. Natl Acad. Sci. USA* **108**, 9449–9454 (2012).
17. Rivera-Cancel, G., Motta-Mena, L. B. & Gardner, K. H. Identification of natural and artificial DNA substrates for light-activated LOV-HTH transcription factor EL222. *Biochemistry* **51**, 10024–10034 (2012).
18. Zoltowski, B. D., Motta-Mena, L. B. & Gardner, K. H. Blue light-induced dimerization of a bacterial LOV-HTH DNA-binding protein. *Biochemistry* **52**, 6653–6661 (2013).
19. Reade, A. *et al.* TAE1: a zebrafish-optimized optogenetic gene expression system with fine spatial and temporal control. *Development* **144**, 345–355 (2017).
20. Jayaraman, P. *et al.* Blue light-mediated transcriptional activation and repression of gene expression in bacteria. *Nucleic Acids Res.* **44**, 6994–7005 (2016).
21. Da Silva, N. A. & Srikrishnan, S. Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **12**, 197–214 (2012).
22. Usherenko, S. *et al.* Photo-sensitive degron variants for tuning protein stability by light. *BMC Syst. Biol.* **8**, 128 (2014).
23. van Maris, A. J. A. *et al.* Directed evolution of pyruvate decarboxylase-negative *Saccharomyces cerevisiae*, yielding a C₂-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl. Environ. Microbiol.* **70**, 159–166 (2004).
24. Klein, J. L. C., Olsson, L. & Nielsen, J. Glucose control in *Saccharomyces cerevisiae*: the role of *MIG1* in metabolic functions. *Microbiology* **144**, 13–24 (1998).
25. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31**, 335–341 (2013).
26. Hammer, S. K. & Avalos, J. L. Uncovering the role of branched-chain amino acid transaminases in *Saccharomyces cerevisiae* isobutanol biosynthesis. *Metab. Eng.* **44**, 302–312 (2017).
27. Park, S. H., Kim, S. & Hahn, J. S. Improvement of isobutanol production in *Saccharomyces cerevisiae* by increasing mitochondrial import of pyruvate through mitochondrial pyruvate carrier. *Appl. Microbiol. Biotechnol.* **100**, 7591–7598 (2016).
28. Deisseroth, K. Optogenetics. *Nat. Methods* **8**, 26–29 (2011).
29. Gerhardt, K. P. *et al.* An open-hardware platform for optogenetics and photobiology. *Sci. Rep.* **6**, 35363 (2016).
30. Miliadis-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P. & Khammash, M. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat. Commun.* **7**, 12546 (2016).
31. Milne, N., Wahl, S. A., van Maris, A. J. A., Pronk, J. T. & Daran, J. M. Excessive by-product formation: a key contributor to low isobutanol yields of engineered *Saccharomyces cerevisiae* strains. *Metab. Eng. Commun.* **3**, 39–51 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Gardner and L. Motta-Mena for providing the plasmids and maps for the EL222 system (pVP16-EL222 and pC120-Fluc)¹², D. Pincus for plasmid pNH603, J. J. Lee for plasmid pET28a Ldh, C. Taxis for plasmid pDS143, C. Nelson for sharing her qPCR equipment, S. Han for assistance in qPCR experiments and figure presentation, J. Rabinowitz and J. Storey for sharing their 500-ml Sixfors fermentation system for fed-batch fermentation experiments, S. Silverman for technical assistance on this equipment and C. DeCoste and the Princeton Molecular Biology Flow Cytometry Resource Center for assistance in flow cytometry experiments. This work was supported by the Alfred P. Sloan Foundation (to J.L.A.), The Pew Charitable Trusts (to J.L.A.), National Institutes of Health grant DP2EB024247 (to J.E.T.) and an Eric and Wendy Schmidt Transformative Technology Fund grant (to J.L.A. and J.E.T.).

Author Contributions E.M.Z., J.E.T. and J.L.A. conceived this project and designed the experiments. E.M.Z., Y.Z. and J.L.A. constructed the strains and plasmids. E.M.Z. and H.P. performed the experiments that are shown in Fig. 1; E.M.Z. and J.M. conducted experiments illustrated in Fig. 2; E.M.Z. performed experiments that are shown in Fig. 3; E.M.Z. performed experiments illustrated in Extended Data Figs 1–10. Y.Z. performed experiments illustrated in Extended Data Figs 1, 8. M.A.L. performed experiments illustrated in Extended Data Fig. 10. E.M.Z., J.E.T. and J.L.A. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.L.A. (javalos@princeton.edu) or J.E.T. (toettcher@princeton.edu).

Reviewer Information Nature thanks L. Larrondo and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Assembly of DNA constructs. We cloned promoter-gene-terminator sequences into previously described standardized vector series (pJLA vectors)²⁵. This allows for easy manipulation and generation of multi-gene plasmids. All genes were designed to have *NheI* and *XhoI* restriction sites at the 5' and 3' ends, respectively, which were used to insert the genes into pJLA vectors. Each promoter-gene-terminator construct is flanked by *XmaI* and *AgeI* restriction sites at their 5' ends, and *MreI*, *AscI* and *BspEI* sites at their 3' ends, which we used for easy assembly of multi-gene plasmids, as previously described²⁵ (Supplementary Table 1).

Qiagen Miniprep, Qiagen Gel Extraction and Qiagen PCR purification kits were used to extract and purify plasmids and DNA fragments. Most genes and promoters (*ILV2*, *ILV3*, *ILV5*, *ARO10*, *LIAdhA*^{RE1} (ref. 32), *GAL4*, *GAL80*, GFP, *P_{GAL1}*, *P_{TEF}*, *P_{TDH3}*, *P_{PGK1}*, *P_{CYC1}*, *P_{ADH1}*) were amplified from yeast genomic DNA or laboratory plasmids, using Phusion polymerase from NEB, according to the manufacturer's instructions. Other genes were amplified from plasmids provided by other groups: *PsLdh* from plasmid pET28a, *Ldh*^{33,34} from J. J. Lee; and the photosensitive degenon derived from the fusion of the LOV2–V19L domain of phototropin 1 from *Arabidopsis thaliana* and a synthetic degradation sequence derived from the murine (mouse) ornithine decarboxylase from plasmid pDS143 from C. Taxis²². The codon-optimized sequence for VP16–EL222 was purchased as a gBlock from IDT. The sequence for the *P_{C120}* promoter was synthesized by the Bio Basic gene synthesis service. When pJLA vectors were not available, we used Gibson isothermal assembly to produce our constructs, based on published protocols³⁵. Enzymes were purchased from NEB (*XmaI*, *AscI*, *NheI*, *XhoI*, *BspEI*, *AgeI*, T4 DNA ligase, Phusion Polymerase) and Thermo Fisher Scientific (*MreI*).

We modified the single copy integration plasmid pNH603³⁶ to make a plasmid compatible with the pJLA vectors that can be used to introduce gene cassettes into the *HIS3* locus. We first removed the *AscI* site of pNH603 and replaced the *ADH1* terminator (*T_{ADH1}*) sequence between *PtsI* and *SacI* with a fragment containing an *XmaI* restriction site. Then we introduced a cloning sequence array consisting of *XmaI*, *MreI*, *AscI* and *BspEI* between the *XmaI* site (which replaced the *T_{ADH1}*) and *KpnI* restriction sites, to make pYZ12-B (Extended Data Fig. 1a, Supplementary Table 1). This addition makes pYZ12-B compatible with the pJLA platform of vectors, and allowed for easy transfer of gene cassettes from pJLA 2μ plasmids. Gene constructs in pYZ12-B were integrated into the *HIS3* locus of the genome by linearizing the plasmid with *PmeI*. We also used pRSII416³⁷ to introduce genes in a single copy episomal plasmid (CEN); this was done by inserting promoter–gene–terminator constructs cut from pJLA vectors with *XmaI* and *MreI* and inserting at the *XmaI* site of pRSII416.

Similar to pYZ12-B, we developed pYZ23 (Supplementary Table 1, Extended Data Fig. 1b), a pJLA vector-compatible plasmid, to integrate multiple copies of gene constructs into the δ -sites of the yeast genome. The pYZ23 plasmid targets the YARCdelta5, the 337-bp-long terminal repeat (LTR) of *S. cerevisiae* Ty1 retrotransposons (YARCTy1-1, SGD ID: S000006792)³⁸. We constructed pYZ23 with four overlapping DNA fragments using the Gibson isothermal assembly method³⁵. The four fragments are: (1) the *PmeI*-linearized backbone fragment from pYZ12-B containing the ampicillin-resistance gene; (2) the first 207 bp of the YARCdelta5 LTR; and (3) the last 218 bp of the YARCdelta5 LTR, both of which were amplified from the BY4741 genome using primer pairs *Yfz_Oli39* and *Yfz_Oli40*, and *Yfz_Oli43* and *Yfz_Oli44*, respectively (Supplementary Table 7); and (4) the *BleMX6* gene cassette from pCY 3090-07 (Addgene 36232), amplified using primers *Yfz_Oli41* and *Yfz_Oli42* (Supplementary Table 7), which add flanking *loxP* sites (*lox66* and *lox71*) to the *BleMX6* gene. Additional restriction sites, including *MreI*, *AscI* and *BspEI*, were introduced for subcloning (Extended Data Fig. 1b, Supplementary Table 7).

All vectors were sequenced using Sanger sequencing from GENEWIZ before using them to transform yeast. We avoid using tandem repeats to prevent recombination after yeast transformation (Extended Data Fig. 1c), and thus do not observe instability of plasmids.

Yeast transformations. Yeast transformations were carried out using standard lithium acetate protocols, and the resulting strains are catalogued in Supplementary Table 2. Gene constructs derived from pYZ12-B and pYZ23 vectors were genomically integrated into the *HIS3* locus and δ -sites (YARCdelta5), respectively. These vectors were first linearized with *PmeI* and DNA fragments were purified using a Qiagen PCR purification kit before using them for yeast transformation. Gene deletions were carried out using homologous recombination strategies. DNA fragments containing antibiotic resistance cassettes flanked with *loxP* sites were amplified using PCR from pAG26³⁹ (containing the hygromycin resistance gene, hygromycin B phosphotransferase (*HygB-PT*)), pUG6⁴⁰ (containing the G418 resistance gene *KanMX*), or pAG36³⁹ (containing the nourseothricin resistance gene *NAT1*), using primers with 40 bp of homology to the promoter and terminator regions of the gene targeted for deletion. Antibiotic-resistance markers were subsequently removed by expressing Cre recombinase from the pSH62 (AF298785)

vector⁴¹. After transformation, cells were plated on synthetic complete (SC) drop out medium depending on the auxotrophy restored by the construct. In the case of antibiotic selection, cells were plated onto non-selective YPD plates for 16h, then replica plated onto YPD plates with 300 μg ml⁻¹ hygromycin (purchased from Invitrogen), 200 μg ml⁻¹ nourseothricin (purchased from WERNER BioAgents), or 200 μg ml⁻¹ G418, purchased from Gibco by Life Technologies). Zeocin was used to select for δ -integration at concentrations ranging from 800 to 1,200 μg ml⁻¹ (purchased from Thermo Fisher Scientific).

All strains with genomic integrations or gene deletions were genotyped with PCR to confirm their accuracy. We integrated constructs in the *HIS3* locus or δ -sites to promote strain stability.

Yeast cell culture growth, centrifugation and optical measurements. Unless otherwise specified, liquid yeast cultures were grown in 24-well plates, at 30 °C and shaken at 200 r.p.m., in either YPD or SC-dropout medium supplemented with 2% glucose. SC-dropout media include 94.8 mg l⁻¹ of valine, 94.8 mg l⁻¹ of isoleucine and 189.6 mg l⁻¹ of leucine along with all other necessary standard nutrients unless otherwise specified. To stimulate cells with light, we used blue light-emitting diode (LED) panels (HQR New Square 12-inch Grow Light Blue LED 14W), placed 40 cm from cell cultures. To control light duty cycles, the LED panels were regulated with a Nearpow Multifunctional Infinite Loop Programmable Plug-in Digital Timer Switch (purchased from Amazon). To resuspend cells in new media, cell cultures were centrifuged in a table-top centrifuge, with 24-well plate rotor adaptors. Unless otherwise specified, plates were centrifuged at 234g for 5 min.

Fluorescence and OD₆₀₀ measurements were taken using a TECAN plate reader (infinite M200PRO). The excitation and emission wavelengths used for GFP fluorescence measurements were 485 nm and 535 nm, respectively, using an optimal gain for all measurements. To process fluorescence data, the background fluorescence from the medium was first subtracted from values. Then, the GFP/OD₆₀₀ values of cells lacking a GFP construct were subtracted from the fluorescence values (GFP/OD₆₀₀) of each sample to normalize for light bleaching of the medium and cell contents. Thus, reported values were calculated according to the following formula

$$\text{GFP/OD}_{\text{strain, condition}} = \frac{\text{GFP}_{\text{strain, condition}} - \text{GFP}_{\text{media, condition}}}{\text{OD}_{\text{strain, condition}} - \text{OD}_{\text{media, condition}}} - \frac{\text{GFP}_{\text{no GFP control strain, condition}} - \text{GFP}_{\text{media, condition}}}{\text{OD}_{\text{no GFP control strain, condition}} - \text{OD}_{\text{media, condition}}}$$

All fluorescence measurements were done at the end of experiments or on samples taken from experimental cultures, such that potential activation of VP16–EL222 by the light used to excite GFP did not affect our experiments or results.

To measure cell concentration, optical-density measurements were taken at 600 nm, using medium (exposed to the same conditions as the yeast) as a blank. Measurements were performed using the TECAN plate reader (infinite M200PRO) or Eppendorf spectrophotometer (BioSpectrometer basic), from samples diluted to a range of OD₆₀₀ of 0.1 to 1.0.

All experiments involving light-inducible strains were done under minimal ambient light, unless otherwise specified, to avoid unwanted activation of optogenetic systems.

Construction of OptoEXP system. We purchased a gBlock (IDT) containing the yeast codon-optimized sequence of VP16–EL222, flanked by *NheI* and *XhoI* restriction sites, which we inserted into plasmid pJLA121⁰³⁰¹ (containing *P_{PGK1}* and *T_{CYC1}*). We then used *XmaI* and *AscI* to subclone *P_{PGK1}*–VP16–EL222–*T_{CYC1}* into pYZ12-B to make EZ-L105 (Supplementary Table 1), which allows single genomic integration of gene cassettes into the *HIS3* locus. We then changed the promoter to *P_{TEF1}* using *XmaI* and *NheI* restriction site cutting and subsequent ligation to make EZ-158, used for OptoEXP expression in glycerol media. In addition, we synthesized the *C120* and minimal promoter sequence (TAGAGGGTATATAATGGAAGCTCGACTTCCAG), otherwise known as *P_{C120}*, using Bio Basic's gene synthesis service and used it to develop new pJLA vectors with the *P_{C120}* promoter and either an *ADH1* or *ACT1* terminator, making pJLA121⁰⁸⁰³ or pJLA121⁰⁸⁰², respectively (Supplementary Table 1).

Characterization of OptoEXP. To test the OptoEXP system, we built plasmid EZ-L83 (pJLA111–GFP⁰⁸⁰³), which places GFP under *P_{C120}* transcriptional control in a CEN/ARS plasmid with a *URA3* marker (Supplementary Table 1). We then used EZ-L105 to integrate a single copy of *P_{PGK1}*–VP16–EL222–*T_{CYC1}* construct into the *HIS3* locus of CENPK.2-1C, selecting strain YE224 from a SC lacking His (SC-His) plate supplemented with 2% glucose. Subsequently, we transformed YE224 with EZ-L83, and selected strain YE232 from a SC plate lacking uracil (SC-Ura) plate supplemented with 2% glucose. We also transformed YE224 with empty pRSII416 to make control strain YE232C, which has no GFP.

To characterize light induction by OptoEXP, we tested four different colonies of YE232. We grew cells from each colony in liquid SC-Ura overnight in the dark

(tin foiled). The next morning, we diluted the cells to OD₆₀₀ of 0.1 in fresh SC-Ura, to a final volume of 1 ml each, incubated in individual wells of a 24-well clear Costar plate. We prepared five identical plates and tin foiled one of them. The plates were then placed in the dark (tin-foil-covered plate), under constant blue light, or under blue light with duty cycles of 5 s on and 75 s off, 8 s on and 72 s off and 11 s on and 69 s off, with a blue LED panel at 40 cm from each plate. We used duty cycles instead of light intensity to better control light dose and reproducibility (Extended Data Fig. 2). Cell cultures were grown for 8 h under blue light panels or in the dark, then their GFP fluorescence and OD₆₀₀ were measured using the TECAN plate reader. Error bars represent the s.d. from biological replicates ($n = 4$).

To benchmark the combination of P_{PGK1}-VP16-EL222-T_{CYC1} and a single copy of P_{C120}, which we refer to as OptoEXP, we compared it to several constitutive promoters. To achieve this, we made pJLA111-GFP^{OXOX} constructs (CEN/ARS with URA3 markers) containing P_{CYC1} (EZ-L64), P_{ADH1} (EZ-L63), P_{PGK1} (EZ-L67), P_{TDH3} (EZ-L65) and P_{TEF1} (EZ-L66), which we used to transform YE224 to make yeast strains YE227 (EZ-L64), YE228 (EZ-L63), YE229 (EZ-L67), YE230 (EZ-L65) and YE231 (EZ-L66).

YE232C was used as a control with no GFP production for strains with pJLA111-GFP^{OXOX} plasmids. YE227, YE228, YE229, YE230, YE231 and YE232 were tested similarly to the OptoEXP testing above (YE232).

Flow cytometry experiments. We used flow cytometry to determine whether the OptoEXP system produces a homogeneous response in the cell population. To construct strains for these experiments, we transformed CEN.PK2-1C with PmeI-linearized pY12-B (a non-fluorescent control), EZ-L136, EZ-L516 or EZ-L350 to make YE2140 (CEN.PK2-1C with histidine prototrophy restored), YE2139 (CEN.PK2-1C with OptoEXP driving GFP, and P_{PGK1} driving VP16-EL222, which works best in glucose), YE2243 (CEN.PK2-1C with OptoEXP driving GFP, and P_{TEF1} driving VP16-EL222, which works well in glucose or glycerol) or YE2186 (CEN.PK-2C with *HIS3::P_{TEF1}-GFP-T_{ACT1}*, for constitutive expression of GFP as control), respectively.

To test the homogeneity of gene expression of OptoEXP we grew overnight cultures of YE2139, YE2140 and YE2186 in SC-His medium supplemented with 2% glucose in the dark. The next morning, we diluted 20 μ l of these cultures into 980 μ l of fresh medium in two 24-well plates. Both plates were placed in the dark and shaken at 200 r.p.m. at 30 °C for 3 h. Then, one plate was placed 40 cm below a blue light panel and the other was tin foiled and kept in the dark. Both plates were shaken at 200 r.p.m. at 30 °C for 3 h. Then, 5 μ l of culture was diluted into 995 μ l of PBS and used for flow cytometry (Extended Data Fig. 2a, b). Samples were run in triplicates from three different cultures separated after the overnight stage. Representative samples from these triplicates were chosen for the figures.

To test OptoEXP performance in non-fermentative conditions, we grew 10-ml overnight cultures of YE2243, YE2140 and YE2186 in SC-His medium supplemented with 3% glycerol and 2% ethanol, in tin-foiled tubes (in the dark) in the roller drum. The next morning, we diluted 30 μ l of these cultures into 980 μ l of fresh medium to an OD₆₀₀ of 0.05–0.1 in four 24-well plates. All four plates were placed in the dark (tin foiled) and shaken at 200 r.p.m. at 30 °C for 6 h (until the yeasts were in the exponential growth phase, OD₆₀₀ of approximately 3). Then, one plate was placed 40 cm below a blue light panel and the others were kept tin foiled (in the dark). After 1 h, another plate was moved into the light. After another half hour more, another plate was moved into the light. All plates were kept shaken at 200 r.p.m. at 30 °C for 2 h after the initial 6 h of growth. Then, 5 μ l of each culture was diluted into 995 μ l of PBS medium and used for flow cytometry (Extended Data Fig. 2c). Samples were run in triplicates from three different cultures separated after the overnight stage. Representative samples from these triplicates were chosen for the figures.

GFP expression was quantified by flow cytometry using a BD LSR II flow cytometer (BD Biosciences) with an excitation wavelength of 488 nm and an emission wavelength of 530 nm. The gating used in our analyses was defined to include positive (YE2186) and negative (YE2140) control cells based on GFP fluorescence, but exclude particles that are either too small or too large to be single living yeast cells, based on the side scatter (SSC-A) versus forward scatter (FSC-A) plots (Extended Data Fig. 2d). Mean fluorescence values were determined from 20,000 cells. Data were analysed with the FlowJo v.10 software (Tree Star).

All fluorescence measurements were performed at the end of experiments or on samples taken from experimental cultures, such that potential activation of VP16-EL222 by the light used to excite GFP did not affect our experiments or results.

Construction of the OptoINVRT circuits. OptoINVRT circuits were initially developed and characterized in two yeast strain backgrounds: YE244 (CENPK2-1C, *GAL80 Δ* , *GAL4 Δ*) and Y202 (S288C, *PDC1 Δ* , *PDC5 Δ* , *PDC6 Δ* , *GAL80 Δ* , containing pJLA121-PDC1⁰²⁰²), (Supplementary Table 2).

Gene circuits were assembled using restriction enzyme digests and ligations afforded by pJLA vectors, in which open reading frames were inserted using NheI and XhoI sites, and multiple cassettes assembled using XmaI (or AgeI), MreI

(or BspEI) and AscI²⁵. Each gene circuit was constructed by assembling five promoter–gene–terminator sequences into single integration vectors targeting the *HIS3* locus (EZ-L259, EZ-L260 and EZ-L266; Supplementary Tables 1, 3). The PSD used in OptoINVRT3 is derived from the fusion of phototropin1 LOV2 domain (V19L mutant) from *A. thaliana* and a synthetic degradation sequence derived from the murine ornithine decarboxylase²². This PSD was fused to the C terminus of *GAL4* using Gibson assembly of pJLA121⁰³⁰³ cut with XhoI and PSD amplified from pDS143²² with overhang sequences of 30 bp that are homologous to the multiple cloning sequence and T_{ADH1} of pJLA121⁰³⁰³ to make EZ-L247, to which we then inserted *GAL4* using NheI and XhoI through ligation to make the precursor plasmid of EZ-L266 (EZ-L251).

Characterization of OptoINVRT circuits. The three OptoINVRT circuits, controlling GFP expression, were initially characterized by transforming YE244 to produce YE2100 (OptoINVRT1), YE2101 (OptoINVRT2) and YE2102 (OptoINVRT3), and Y202 to produce YE2115 (OptoINVRT1), YE2116 (OptoINVRT2) and YE2117 (OptoINVRT3) (Supplementary Table 2). OptoINVRT circuits were characterized using strains YE2100–102 and YE2115–117 by performing the same experiment done on YE232, described above, to characterize the OptoEXP system. In this case, we exposed cells to full light, complete darkness, or light pulses of 8 s on and 72 s off. YE2186 and YE2140 were used as controls for strains derived from YE244, and YE2171 and YE294 were used as controls for strains derived from Y202.

Characterization of light panels. The light intensity of panels was measured using a Quantum meter with a separate sensor (Model MQ-510 from Apogee Instruments), placed 40 cm from the light panel of interest (the same distance was used for all 24-well plate experiments). The resulting intensities ranged from 53 to 134 μ mol m⁻² s⁻¹ with an average of 77.5 μ mol m⁻² s⁻¹ (with 465 nm max peak spectra). Ambient light under these conditions ranged from 4–5 μ mol m⁻² s⁻¹. We then measured the levels of GFP expression in strains YE2243 and YE2186 (as well as YE2140, as a negative control) induced by three different panels with 53, 82 or 134 μ mol m⁻² s⁻¹ intensities, placed 40 cm from the 24-well plates, using the same protocol for experiments shown in Fig. 1b. Overnight cultures grown in the dark where diluted to an OD₆₀₀ of 0.1, and incubated for 8 h in either full light, or light pulses of 10 s on and 70 s off. The levels of GFP expression were indistinguishable across the three panels tested (including the brightest and dimmest panels used in this study), and responded uniformly to changes in light duty cycle (Extended Data Fig. 3). This demonstrates that there is no detectable difference between variations in the intensities of panels used in this study and that varying duty cycle is an effective way to control gene expression using our optogenetic transcriptional controls.

Development of an OptoEXP-PDC strain. Strain Y200 contains a triple gene deletion of *PDC1 Δ* , *PDC5 Δ* and *PDC6 Δ* , as well as a 2 μ -URA3 plasmid pJLA121-PDC1⁰²⁰² with P_{TEF1}-PDC1-T_{ACT1}, which allows it to grow robustly in glucose. Y200 was transformed with a PmeI-linearized EZ-L165 plasmid to insert a cassette composed of P_{TEF1}-VP16-EL222-T_{CYC1} and P_{C120}-PDC1-T_{ADH1} into its *HIS3* locus, resulting in strain YE250 (Supplementary Tables 1, 2). As a control, we also transformed Y200 with PmeI-linearized EZ-L158, a vector containing P_{TEF1}-VP16-EL222-T_{CYC1} but lacking P_{C120}-PDC1-T_{ADH1}, and then counter-selected against the pJLA121-PDC1⁰²⁰² plasmid by growing on 5-fluoroorotic acid (5-FOA) (as described below) to produce the control strain YE250C. Strain YE250 was then transformed with PmeI-linearized EZ-L143, which inserts multiple copies of P_{C120}-PDC1-T_{ADH1} into δ -integration sites of the yeast genome. Colonies able to grow on YPD plates containing 800 μ g ml⁻¹ of zeocin were replica plated on plates containing SC-His with 3% glycerol and 2% ethanol twice. The resulting plates were then replica plated on SC-His with 3% glycerol, 2% ethanol and 1 mg ml⁻¹ 5-FOA, and then finally back onto plates containing SC-His with 3% glycerol and 2% ethanol. This treatment efficiently counter-selects against pJLA121-PDC1⁰²⁰², due to its URA3 marker⁴². From this plate, we isolated YE261-23, a strain that can grow on SC-His with 2% glucose plates only when exposed to blue light, which is consistent with having *PDC1* expression controlled by P_{C120}, and VP16-EL222. As a control, we also transformed Y200 with PmeI-linearized EZ-L143, which contains P_{C120}-PDC1-T_{ADH1} but lacks P_{TEF1}-VP16-EL222-T_{CYC1}, and counter-selected against in pJLA121-PDC1⁰²⁰² in 5-FOA to produce YE261C. This control strain has multiple copies of P_{C120}-PDC1-T_{ADH1} in δ -sites, but lacks the VP16-EL222 needed to transcribe them.

Light-dependent growth of an OptoEXP-PDC strain. Solid media. Cells from strains BY4741, YE261-23, YE250C and YE261C were patched onto an agar plate containing yeast extract, peptone and 3% glycerol and 2% ethanol (YPGE plate), and grown over night in ambient light. This plate was then replica plated onto a fresh YPGE plate and two YPD plates. One of the YPD plates was covered in tin foil and the other exposed to constant blue light, while the YPGE was left at ambient lighting. Replica plating was performed such that the YPD plate ultimately covered in tin foil was replicated first, the YPD plate ultimately exposed to constant blue

light was replicated second, and the YPGE plate was replicated last. All plates were incubated at 30 °C for 48 h (Fig. 2b).

Liquid media. Single colonies of YEZ61-23 and BY4741 were used to inoculate liquid SC-His with 2% glucose medium. The cells were grown for 24 h on a roller drum at 200 r.p.m., 30 °C and 40 cm away from a blue light source (HQRP New Square 12-inch LED Grow Light System 225 Blue LED 14W). Subsequently, the OD₆₀₀ was measured with a spectrophotometer, and the cells were diluted to an OD₆₀₀ of 0.1 in fresh SC-His medium supplemented with 2% glucose medium in three 1-ml replicates in four 24-well plates (Celltreat non-treated sterile flat bottom plates). Each plate was centred 40 cm under an HQRP blue light panel. One plate was exposed to full blue light. The other three plates were exposed to 10 s on and 70 s off of blue light, 20 s on and 60 s off of blue light or 40 s on and 40 s off of blue light. A fifth plate was wrapped in aluminium foil to incubate cells in the absence of light. OD readings were taken using a TECAN plate reader at 0 h, 10 h, 12.5 h, 13.5 h, 15 h, 17 h, 19 h, 20.5 h and 31.75 h. Readings were taken under minimal light conditions to prevent unwanted activation of EL222. Data are available in a supplementary spreadsheet.

The exponential growth phase of YEZ61-23 (identified as the most linear portion of the plot of log_e(OD) against time), was used to find the specific growth rates at different light doses. This was done by fitting the data to log_e(OD) = log_e(OD₀) + $\mu \times t$, using least squares linear regression, where OD₀ is a constant corresponding to the initial OD, and μ corresponds to the specific growth rate constant. The μ constants were calculated for each independent experiment and then averaged, with error bars representing s.d. ($n = 3$). Source Data are available with the online version of this paper.

Development of chemical production strains. To develop light-dependent lactic acid-producing strains, we transformed Y202 with PmeI-linearized EZ-L259, EZ-L260 and EZ-L266 (OptoINVRT1, OptoINVRT2 and OptoINVRT3 respectively) yielding YEZ115, YEZ116 and YEZ117. We then integrated multiple copies of PmeI-linearized EZ-L235 (Supplementary Table 1), which contains P_{CI20} driving *PDC1* and P_{GAL1} driving the *Ldh* from *Pelodiscus sinensis* (provided by J. Lee) into δ -sites. Then we counter-selected against pJLA121-PDC1⁰²⁰² to produce YEZ144 (OptoINVRT1), YEZ145 (OptoINVRT2) and YEZ146 (OptoINVRT3) (Supplementary Table 2). These strains induce *PDC1* and repress *Ldh* expression in the light; while in the dark they stop inducing *PDC1* and induce *Ldh* instead.

To develop light-dependent isobutanol-producing strains, we transformed YEZ115, YEZ116 and YEZ117 with PmeI-linearized EZ-L316, which integrates multiple copies of P_{CI20}-driven *PDC1* and P_{GAL1}-driven *ILV2* in genomic δ -integration sites. We then counter-selected the transformants against pJLA121-PDC1⁰²⁰² with 5-FOA to produce strains YEZ131 (OptoINVRT1), YEZ149 (OptoINVRT2) and YEZ133 (OptoINVRT3), respectively. Subsequently, we transformed these strains with plasmid EZ-L310 (Supplementary Table 1), which contains five genes from the mitochondrial isobutanol biosynthetic pathway: *ILV2*, *ILV5*, *ILV3*, *COX4-ARO10* and *COX4-LIAdha*^{RE1} (ref. 25). The last two genes in EZ-L310 are fused by their N termini to the mitochondrial localization signal of *COX4*, ensuring that all five genes are targeted to the mitochondria²⁵. In addition, P_{GAL1} drives *ILV2* expression, which places this gene under the control of the OptoINVRT circuits. The resulting strains are YEZ159, YEZ156 and HPY6 for OptoINVRT1, OptoINVRT2 and OptoINVRT3, respectively (Supplementary Table 2).

Screens for lactic acid- and isobutanol-producing strains. Colonies from each transformation plate (grown in glucose and under blue light) were screened for lactic acid (7 colonies from each of YEZ144, YEZ145 and YEZ146) or isobutanol (12 colonies from each of YEZ159, YEZ156 and HPY6) production (Supplementary Table 2). Each colony was used to inoculate 1 ml SC-His medium supplemented with 2% glucose (for lactic acid-producing strains) or SC-Ura medium supplemented with 2% glucose (for isobutanol-producing strains) in 24-well plates and grown overnight at 30 °C, 200 r.p.m. and blue light. The next morning, each culture was diluted to an OD₆₀₀ of 0.1 for lactic acid-producing strains, or an OD₆₀₀ of 0.15 for isobutanol-producing strains, in fresh medium, and grown for 12 h (for lactic acid production) or 18 h (for isobutanol production), at 30 °C, 200 r.p.m. and under blue light. After these incubation periods, the cultures reached OD₆₀₀ values of 5 (for lactic acid-producing strains, $\rho = 5$) and 5 and 8 (for isobutanol-producing strains, $\rho = 8$, $\rho = 5$); at which point they were moved into the dark for 6 h for lactic acid-producing strains ($\theta = 6$ h) and 3 h for isobutanol-producing strains ($\theta = 3$ h). The cultures were then centrifuged and re-suspended in fresh medium, in plates that were then sealed with Nunc Sealing Tape (Thermo Fisher Scientific) to begin the fermentations. The plates were incubated in the dark at 30 °C and shaken at 200 r.p.m. for 48 h during fermentation. Subsequently, the cultures were centrifuged, and the supernatants collected for high-performance liquid chromatography (HPLC) analysis. For lactic acid-producing strains, most of the seven colonies screened for each OptoINVRT circuit produced similar amounts of lactic acid, with colonies of strain YEZ145 (containing OptoINVRT2) producing the

highest titres (Extended Data Fig. 4c). On the other hand, out of the 12 colonies screened for each isobutanol-producing strain, only 2 or 3 produced high titres for each OptoINVRT circuit, with YEZ159 (containing OptoINVRT1) producing the highest titres (Extended Data Fig. 4d, e).

Construction of a high isobutanol-producing strain. We deleted *BAT1* from YEZ131 via homologous recombination, using the HygB hygromycin-resistance marker, resulting in strain YEZ158. Subsequently, we transformed YEZ158 with EZ-L310, resulting in transformants YEZ167, from which we screened 7 colonies, as described above, and identified YEZ167-4 as the strain with highest isobutanol production.

Optimizing experimental parameters for light-dependent fermentation. The colonies producing the highest titres from YEZ144, YEZ145 and YEZ146 (for lactic acid) and of YEZ167 (specifically YEZ167-4 for isobutanol), were used to optimize the pre-growth parameters of fermentation for lactic acid or isobutanol production. For each strain, an overnight culture was grown in blue light at 30 °C with shaking at 200 r.p.m., in SC medium supplemented with 2% glucose (SC-His for lactic acid-producing strains and SC-Ura for isobutanol-producing strains). To find the optimal cell density at which to switch cultures from light to dark, we diluted the overnight cultures into 1 ml SC-dropout medium to different OD₆₀₀ values, ranging from 0.04 to 0.32. The lactic acid-producing strains were then grown for 16 h under 15 s on and 65 s off blue light cycles. The isobutanol-producing strains were grown for 18 h under 15 s on and 65 s off blue light cycles. We then measured the OD₆₀₀ of each culture (these values correspond to variations in ρ), and incubated them in the dark for 6 h for lactic acid-producing strains ($\theta = 6$ h) and 3 h for isobutanol-producing strains ($\theta = 3$ h). After this dark incubation period, the cultures were centrifuged at 234g for 5 min and suspended in fresh SC-dropout medium containing glucose at 26.5 g l⁻¹ (for lactic acid-producing strains) or 21.5 g l⁻¹ (for isobutanol-producing strains). The plates were sealed with Nunc Sealing Tape, and incubated in the dark for fermentation at 30 °C and 200 r.p.m. Control cultures were grown under 15 s on and 65 s off blue light during the growth phase, the dark incubation period, and the fermentation. Cultures producing lactic acid were harvested after 48 h, while samples of cultures producing isobutanol were taken after 24, 48 and 72 h. Cultures were centrifuged at 234g for 10 min, and supernatants were analysed with HPLC.

To optimize the dark incubation period immediately before fermentation (θ), the best isobutanol-producing strain, YEZ167-4, was grown overnight under 15 s on 65 s off blue light in SC-Ura medium supplemented with 2% glucose. The overnight culture was then diluted into seven different plates in quadruplicate samples in fresh medium to a starting OD₆₀₀ of 0.1. The cultures were then grown to an OD₆₀₀ of 8.5 (which was found to be the optimal OD₆₀₀ in our previous experiment). At that point, the plates were tin foiled to ensure complete darkness. After every hour, one of the plates was centrifuged, and the cells suspended in fresh SC-Ura medium with 20.8 g l⁻¹ glucose medium and subjected to a 48-h fermentation in the dark.

Optimizing the frequency of light bouts during the production phase of batch fermentation. A single colony of the best isobutanol-producing strain, YEZ167-4, was used to inoculate 5 ml SC-Ura medium supplemented with 4% glucose and grown overnight under light. The next morning, the culture was diluted in 1 ml fresh medium to an OD₆₀₀ of 0.2 (in quadruplicates) and grown under light for 20 h to an OD₆₀₀ of 9.5. Subsequently, the cultures were incubated for 3 h in the dark. To start the fermentations, the cultures were centrifuged again, and suspended in fresh SC-Ura medium supplemented with 15% glucose (precisely 157.0 g l⁻¹ glucose, as measured with HPLC), and kept in the dark. During the fermentation, the cultures were first exposed to 4 h of light and then pulsed every 5, 10, or 20 h for 30 min, at a duty cycle of 15 s on and 65 s off. As controls, some plates were always kept in the dark or in full light. Fermentations lasted for 80 h, after which the cultures were centrifuged, and the supernatants were analysed with HPLC.

NAD⁺/NADH ratio measurements. YEZ167-4 was inoculated into 3 ml SC-Ura medium supplemented with 2% glucose and grown overnight under light at 30 °C and 200 r.p.m. The next morning, the culture was diluted into 1 ml fresh medium in 6 different 24-well plates to an OD₆₀₀ of 0.2 (in quadruplicates) and grown at 30 °C and shaken at 200 r.p.m. under light for 20 h to an OD₆₀₀ of 9.5. Subsequently, the cultures were incubated under the same conditions for 3 h in the dark. The cultures were then centrifuged at 234g for 5 min and resuspended in fresh SC-Ura medium supplemented with 15% glucose (precisely 157.0 g l⁻¹ glucose, as measured with HPLC), and kept in the dark. As controls, one plate was always kept in the dark and one plate was kept in full light. The remaining four plates were pulsed every 10 h for 30 min, at a duty cycle of 15 s on and 65 s off. One plate was exposed to an extra 30-min pulse of light at a duty cycle of 15 s on and 65 s off precisely 4 h before harvest and placed back into the dark for the remainder of the fermentation. One plate was exposed to an extra 30-min pulse of light at

a duty cycle of 15 s on and 65 s off precisely 30 min before harvesting to test how the light pulses affect the NAD^+/NADH ratio. After 48 h, cells were harvested by centrifuging at 9,000g for 5 min and removing all supernatant. The NAD^+/NADH ratios of the pelleted cells were measured as previously described⁴³.

Light-dependent growth in a 2-l bioreactor. To test whether light penetration becomes limiting to a strain with light-dependent growth, we used a 2-l photobioreactor to compare the growth of YEZ167-4 (which is an isobutanol-producing strain with light-dependent growth) to that of YZY335 (which is a strain that constitutively produces isobutanol and is PDC^+ , and thus its growth and isobutanol production are independent of light).

We used a single colony of YEZ167-4 or YZY335 to inoculate 5 ml SC-Ura medium supplemented with 2% glucose under light, overnight. The next morning, we diluted the culture in 0.5 l SC-Ura medium supplemented with 15% glucose to an OD_{600} of 0.1 in a UTEX 2-l glass photobioreactor surrounded by three blue light panels, placed 1 cm from its glass wall (Extended Data Fig. 9a, b), instead of the UTEX LED lighting platform. The culture was grown for 84 h at 30 °C and under constant blue illumination, with cells gently mixed with a magnetic stir bar and air sparging. Samples were taken every 12 h to measure the OD_{600} of the cell cultures (Fig. 3d). We report the average measurements of three independent fermentations for each strain, and the error bars correspond to the standard deviations of those three measurements.

Fed-batch fermentation. To test the capability of these systems to utilize higher amounts of glucose in a laboratory-scale fermenter, we conducted fed-batch experiments using the Sixfors INFORS AG CH-4103 Boltmengen/Switzerland based on a previously described protocol⁴⁴. We did not sparge air to keep the fermentation in microaerobic conditions. The pH was set to 5.5 and maintained with a 0.5 M KOH feed. We autoclaved the 500-ml fermenter with 250 ml ddH_2O and exchanged the ddH_2O (using the air pump) with filtered 250 ml SC-Ura medium supplemented with 10% glucose. We used a single colony of YEZ167-4 to inoculate 5 ml SC-Ura medium supplemented with 2% glucose under light, overnight. The next day, we added the inoculum to the fermenter to obtain an OD_{600} of 0.2. The fermentation was kept at 30 °C and mixed at 200 r.p.m. Two light panels were placed 0.5 cm away from the vessel walls (Extended Data Fig. 9c, d), and kept on for the first 46 h (growth phase of the fermentation), in which the culture grew to an OD_{600} of 8.2 ± 0.1 . The culture was then covered with black cloth and the light was turned off for 4 h. We then started a glucose feed of SC-Ura medium supplemented with 50% glucose at a rate of 2.9 ml h^{-1} for 90 h (run 1) or 120 h (runs 2 and 3) and turned on the light (continuously) for the first 4 h of the glucose feed. After the initial 4 h, the light was switched to a duty cycle of 45 min on and 7 h 15 min off for the remainder of the fermentation (production phase). Samples of 1 ml were taken at hours 0, 24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 264 (after inoculation), for both OD_{600} measurements, and HPLC analysis. Samples collected at 48, 96, 144, 192 and 264 h were also analysed with qPCR.

Analysis of chemical concentrations. The concentrations of glucose, lactic acid, ethanol, isobutanol and 2-methyl-1-butanol were quantified with HPLC, using an Agilent 1260 Infinity instrument (Agilent Technologies). Samples were centrifuged to remove cells and other solid debris at 234g for 10 min, and the supernatants were then centrifuged for 30 min at 17,000g and 4 °C prior to loading onto an Aminex HPLC-87H ion-exchange column (Bio-Rad). The column was eluted with a mobile phase of 5 mM sulfuric acid at 55 °C and a flow rate of 0.6 ml min^{-1} . Glucose, lactic acid, ethanol, isobutanol, and 2-methyl-1-butanol were monitored with a refractive index detector (RID). To determine their concentration, the peak areas were measured and compared to those of standard solutions for quantification.

Quantification of gene expression and δ -integration copy number. To measure the number of copies of $\text{P}_{\text{C120}}\text{-PDC1}$ integrated into the genome of YEZ50lost, YEZ61-23, YEZ144, YEZ145, YEZ146 and YEZ167-4, we ran qPCR experiments on their genomic DNA, extracted with phenol-chloroform⁴⁵. As a control strain containing a single copy of PDC1 , we developed YEZ50lost by counter-selecting against plasmid pJLA121- PDC1^{0202} from strain YEZ50, using 5-FOA as described above.

To quantify the expression of PDC1 and ILV2 during the fed-batch fermentation experiments, we measured transcript levels using qPCR. We collected 5 ml samples over the course of the fermentation and centrifuged them at 900g for 1.5 min. The mRNA was then extracted with phenol-chloroform⁴⁵. Concentrations of samples were measured using the absorbance at 260 nm/280 nm. The QuantiTect Reverse Transcription Kit (Qiagen) was used to remove genomic DNA and synthesize cDNA.

qPCR was performed using the Bio-Rad Mini Opticon and the iTaq Universal SYBR Green Supermix (Bio-Rad). Reactions contained the appropriate amounts of

reagents as specified by kit protocols. Threshold cycle values were set by StepOne software (Thermo Fisher Scientific) and melting curve analysis following amplification was used to verify the quantity of single PCR products. Primers specific for PDC1 , ILV2 and ACT1 are listed in Supplementary Table 7. The expression level of each gene was normalized to that of ACT1 from the same samples.

Testing gene expression of OptoEXP in 5-l fermenter and higher cell densities.

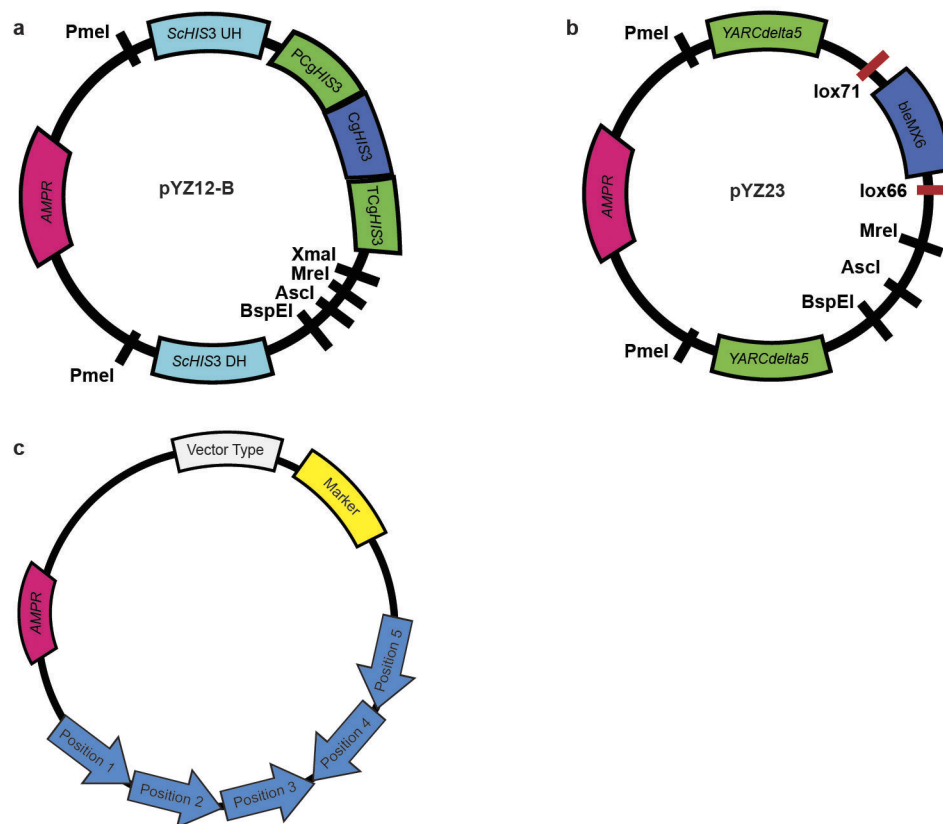
To test OptoEXP in higher cell density conditions, we inoculated YEZ243 into SC-His medium supplemented with 3% glycerol and 2% ethanol and grew in the dark for 24 h. We then set up a BioFlo120 system with a 5-l bioreactor (Eppendorf, B120110002) and added 2.4 l of SC-His medium supplemented with 3% glycerol and 2% ethanol after autoclaving. The reactor was set to 30 °C, pH of 6.5, and a minimum dissolved oxygen percentage of 40. Blue LED strips (Shenzhen Shinesky Optoelectronics, SMD3528) were wrapped around the reactor, covering 73% of the available bulk surface area of the fermentation (Extended Data Fig. 10a, b). These LED strips emitted light at an intensity of $129 \mu\text{moles m}^{-2} \text{ s}^{-1}$ (with 465 nm max peak spectra) measured in the same condition as the light panels. The reactor was then inoculated to an OD_{600} of 1 and the cells were grown in the dark until an OD_{600} of 15 (maintained by covering the reactor with black fabric). At an OD_{600} of 15, the lights were turned on and samples were taken at an OD_{600} of 16 (1 h after induction), 19 (6 h after induction), 41 (24 h after induction), 46 (32 h after induction) and 50 (40 h after induction). Cells were fixed by diluting them to an OD_{600} of 1 in SC-His medium with 3% glycerol and 2% ethanol, adding paraformaldehyde to 3.7%, and incubating for 1 h at 25 °C. To prepare samples for flow cytometry, cells were washed twice with Dulbecco's PBS, and re-suspended again in Dulbecco's PBS to an OD_{600} of 0.5.

Statistics. Statistical significance was determined using a standard t -test for P values. T scores were calculated by the formula:
$$\frac{(\text{mean}_{\text{condition 2}} - \text{mean}_{\text{condition 1}}) \sqrt{\text{number of samples}}}{\text{s.d.}_{\text{condition 2}}}$$

P values were calculated using a degree of freedom of 2 and a one-sided t -test calculator.

Data availability. The authors declare that all data supporting the findings of this study are available within the paper (and its Supplementary Information files), but original data that supports the findings are available from the corresponding authors upon reasonable request.

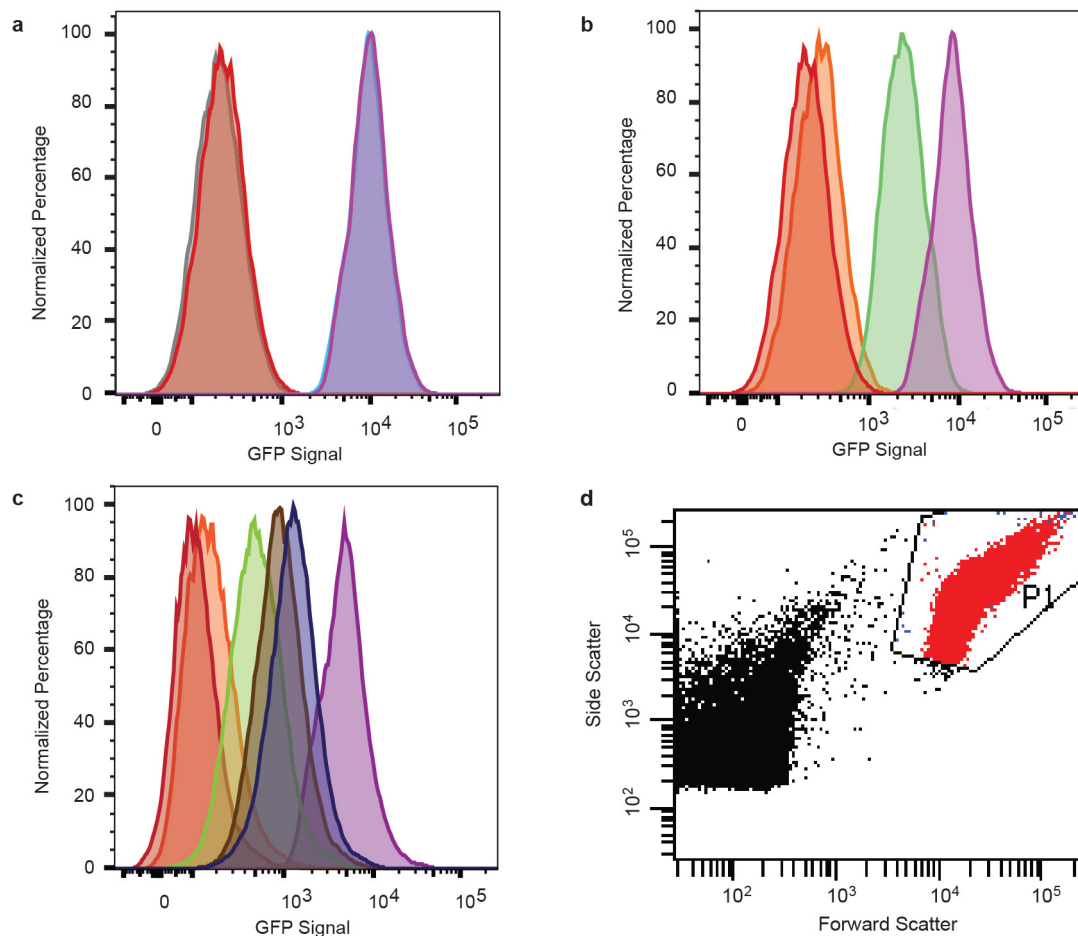
32. Liu, X. *et al.* Structure-guided engineering of *Lactococcus lactis* alcohol dehydrogenase LiAdhA for improved conversion of isobutyraldehyde to isobutanol. *J. Biotechnol.* **164**, 188–195
33. Lee, J. J., Crook, N., Sun, J. & Alper, H. S. Improvement of lactic acid production in *Saccharomyces cerevisiae* by a deletion of *ssb1*. *J. Ind. Microbiol. Biotechnol.* **43**, 87–96 (2016).
34. Lee, J. Y., Kang, C. D., Lee, S. H., Park, Y. K. & Cho, K. M. Engineering cellular redox balance in *Saccharomyces cerevisiae* for improved production of L-lactic acid. *Biotechnol. Bioeng.* **112**, 751–758 (2015).
35. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
36. Youk, H. & Lim, W. A. Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* **343**, 1242782 (2014).
37. Chee, M. K. & Haase, S. B. New and redesigned pRS plasmid shuttle vectors for genetic manipulation of *Saccharomyces cerevisiae*. *G3 (Bethesda)* **2**, 515–526 (2012).
38. Yuan, J. & Ching, C. B. Combinatorial assembly of large biochemical pathways into yeast chromosomes for improved production of value-added compounds. *ACS Synth. Biol.* **4**, 23–31 (2015).
39. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541–1553 (1999).
40. Güldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519–2524 (1996).
41. Güldener, U., Heinisch, J., Koehler, G. J., Voss, D. & Hegemann, J. H. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30**, e23 (2002).
42. Jones, E. W. & Fink, G. R. in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression* (eds Strathern, J. N. *et al.*) 181–299 (Cold Spring Harbor, 1982).
43. Kern, S. E., Price-Whelan, A. & Newman, D. K. Extraction and measurement of NAD(P)^+ and NAD(P)H . *Methods Mol. Biol.* **1149**, 311–323 (2014).
44. Ziv, N., Brandt, N. J. & Gresham, D. The use of chemostats in microbial systems biology. *J. Vis. Exp.* **80**, e50168 (2013).
45. Collart, M. A. & Oliviero, S. Preparation of yeast RNA. *Curr. Protoc. Mol. Biol.* **Chapter 13**, 12 (2001).



Extended Data Figure 1 | Maps of key vectors used in this study.

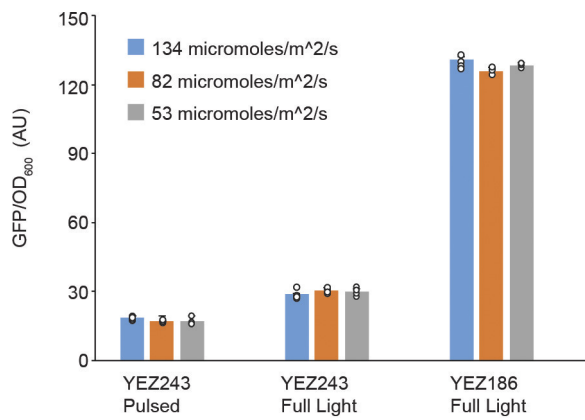
a, pYZ12-B vector used to integrate genes or circuits into the *HIS3* locus. Constructs are usually transferred from pJLA vectors using XmaI and AscI sites. Final constructs are linearized with PmeI before yeast transformation. **b**, pYZ23 vector used to integrate genes or circuits into δ -5 sites of yeast. Constructs are usually transferred from pJLA vectors using MreI and AscI sites. Final constructs are linearized with PmeI before yeast

transformation. **c**, The general vector map shows the relative orientation of the five positions listed in Supplementary Table 1, in which different genes (including promoters and terminators) were assembled, using a previously described multiple gene insertion strategy²⁵. All vectors have an ampicillin-resistance marker (*AMPR*) for cloning in *E. coli* and a selection marker for *S. cerevisiae* (Marker). Vector types include CEN/ARS, 2 μ , or integrative^{36–39}.

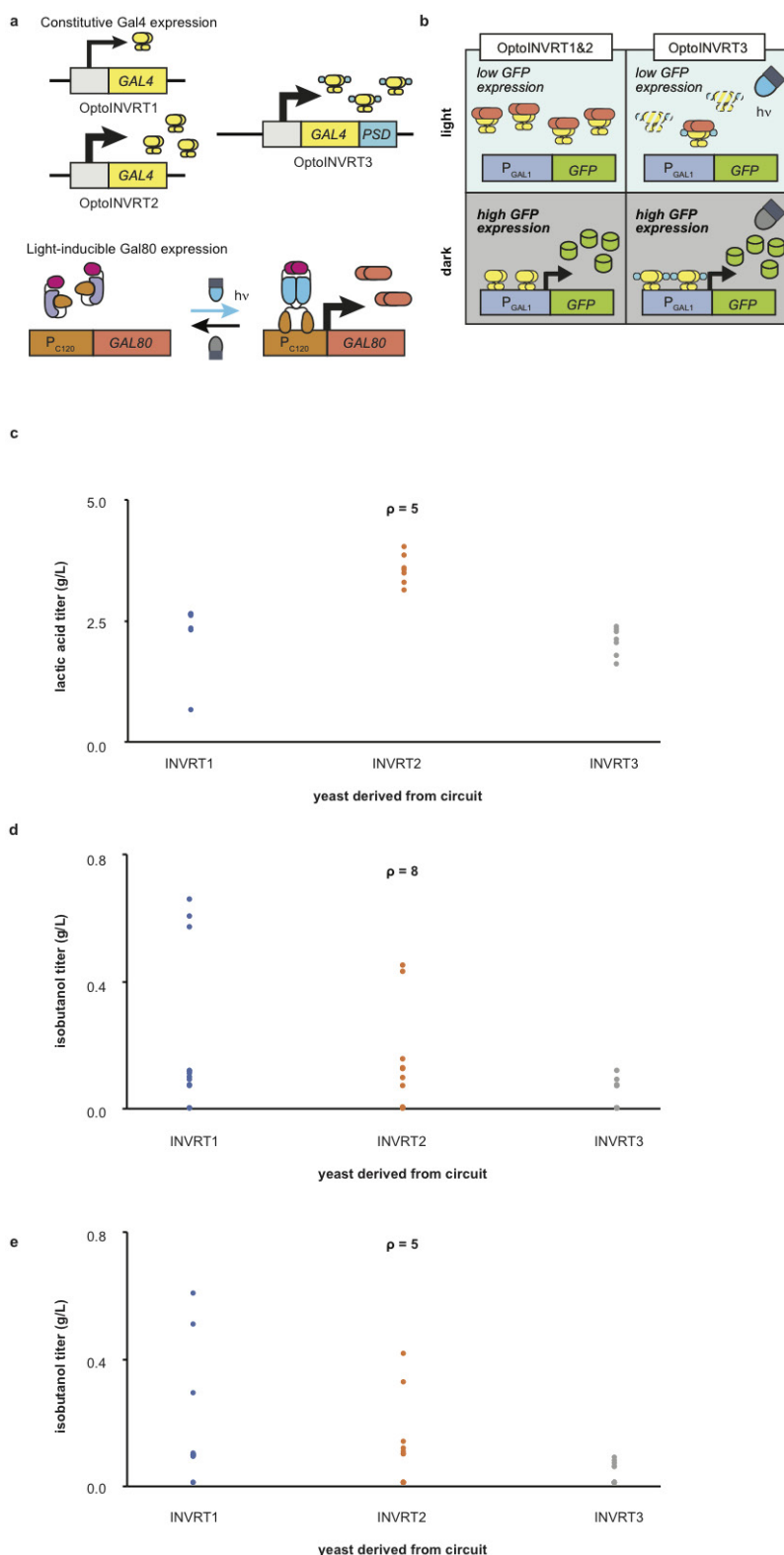


Extended Data Figure 2 | Flow cytometry of strain with GFP controlled by OptoEXP. Representative flow cytometry graphs from three biological replicates under the same conditions. Experiments were performed in 24-well plates, in either glucose or glycerol and ethanol. Every graph is generated from 20,000 cells. **a**, Control strain (YEZ186) with GFP under P_{TEF1} exposed for 3 h to constant blue light (magenta) or kept for the same amount of time in the dark (cyan); these samples are almost completely superimposed on the right-hand side of graph. Control strain (YEZ140) with no GFP exposed for 3 h of constant blue light (red) or kept for the same amount of time in the dark (grey); these samples are almost completely superimposed on the left-hand side of graph. Under these conditions, there is no detectable photobleaching. **b**, Light-induced GFP expression in YEZ139, a strain with GFP controlled by OptoEXP. GFP expression in YEZ139 in SC-His medium supplemented with 2% glucose after 3 h of exposure to blue light (green) is homogeneous across the cell population, and 37-fold higher than in YEZ139 cells kept in the dark for

the same amount of time (orange). The maximum level of GFP expression obtained by OptoEXP in YEZ139 grown in full light for 3 h (green) is 22.1% of what is achieved in YEZ186, which contains P_{TEF1} -GFP, grown under the same conditions (magenta). Fluorescence from a control wild-type strain without GFP, YEZ140, grown for 3 h in the light, is shown for comparison (red). **c**, Light-induced GFP expression by OptoEXP in YEZ243 in SC-His medium supplemented with 3% glycerol and 2% ethanol. Starting from cultures grown in the dark, samples were taken (in the exponential growth phase, at an OD_{600} of approximately 3) and incubated in 24-well plates under the following light conditions: 2 h in the dark (orange); 1.5 h in the dark followed by 30 min in light (green); 1 h in the dark, followed by 1 h in light (brown); or 2 h in light (dark blue). YEZ140 (red) and YEZ186 (magenta) were used as controls with no GFP expression and GFP expression from a strong, constitutive promoter (P_{TEF1}), respectively. **d**, Example of the gating used to make the flow cytometry plots in **a–c**. All experiments were repeated at least three times.

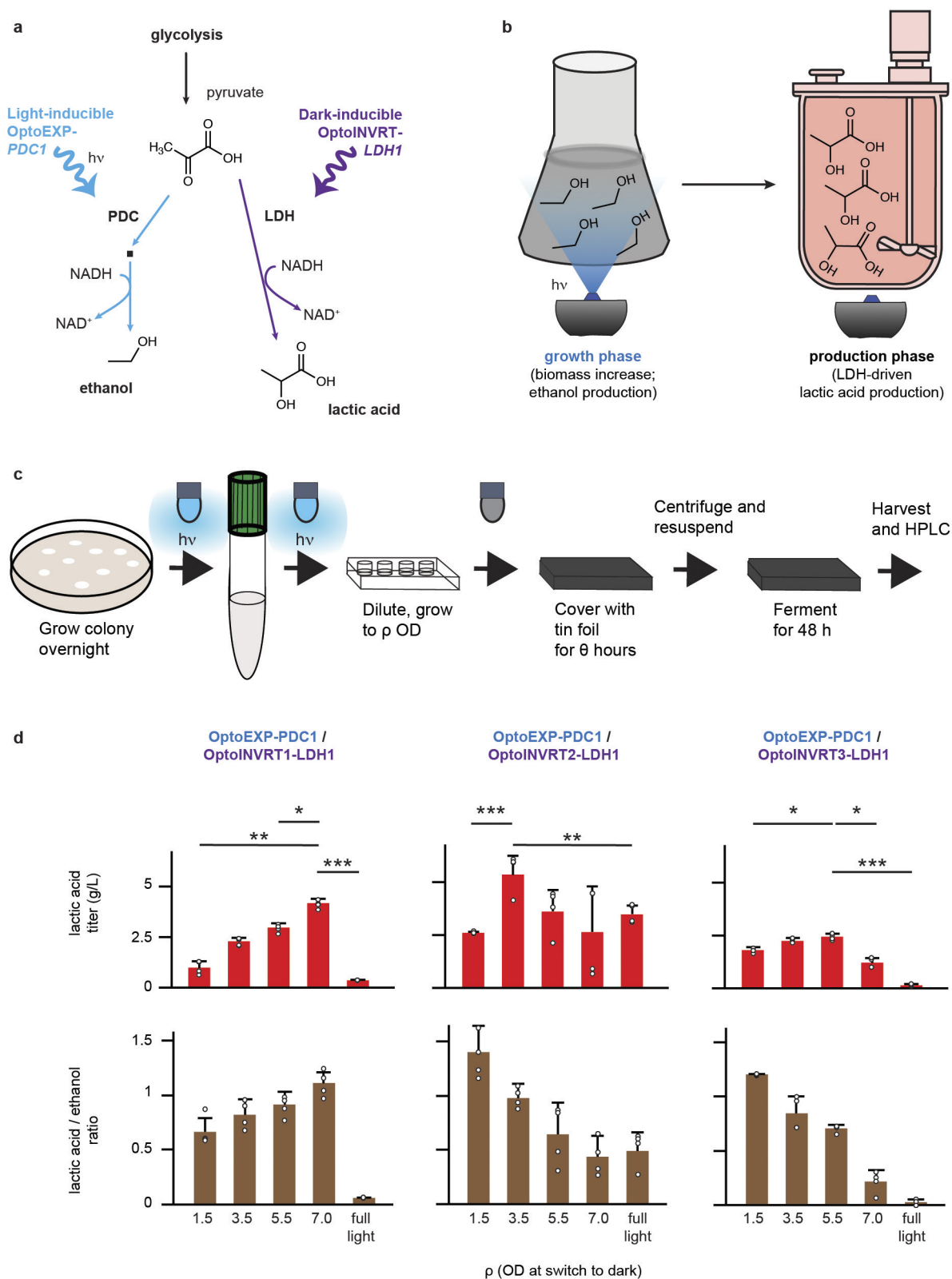


Extended Data Figure 3 | OptoEXP performance under LED panels of different intensities. Light-induced expression of GFP in YEZ243 (OptoEXP driving GFP) compared to constitutive GFP expression in YEZ186 (P_{TEF1} -GFP), using light panels of different intensities. Pulsed light was applied at duty cycles of 10 s on and 70 s off. Data are shown as mean values; dots represent individual data points; error bars represent the s.d. from four biologically independent 1-ml culture sample replicates. All experiments were repeated at least three times.



Extended Data Figure 4 | Comparison of OptoINVRT light-repressible transcription circuits. **a**, OptoINVRT circuit design, based on the expression of Gal80 from OptoEXP and of Gal4 from constitutive promoters of different strength, with or without a PSD domain. **b**, Genes controlled by OptoINVRT circuits are repressed in the light and activated in the dark by the repression activity of Gal80 on Gal4 transcription factor. The PSD fused to Gal4 in OptoINVRT3 stimulates protein degradation in the light. **c**, Screens for lactic acid production in 2% glucose of several colonies of strains YEZ144 (OptoINVRT1), YEZ145 (OptoINVRT2) and YEZ146 (OptoINVRT3), using growth parameters: $\rho = 5$ and $\theta = 6$ h, where ρ is the cell density at which cells are moved from light

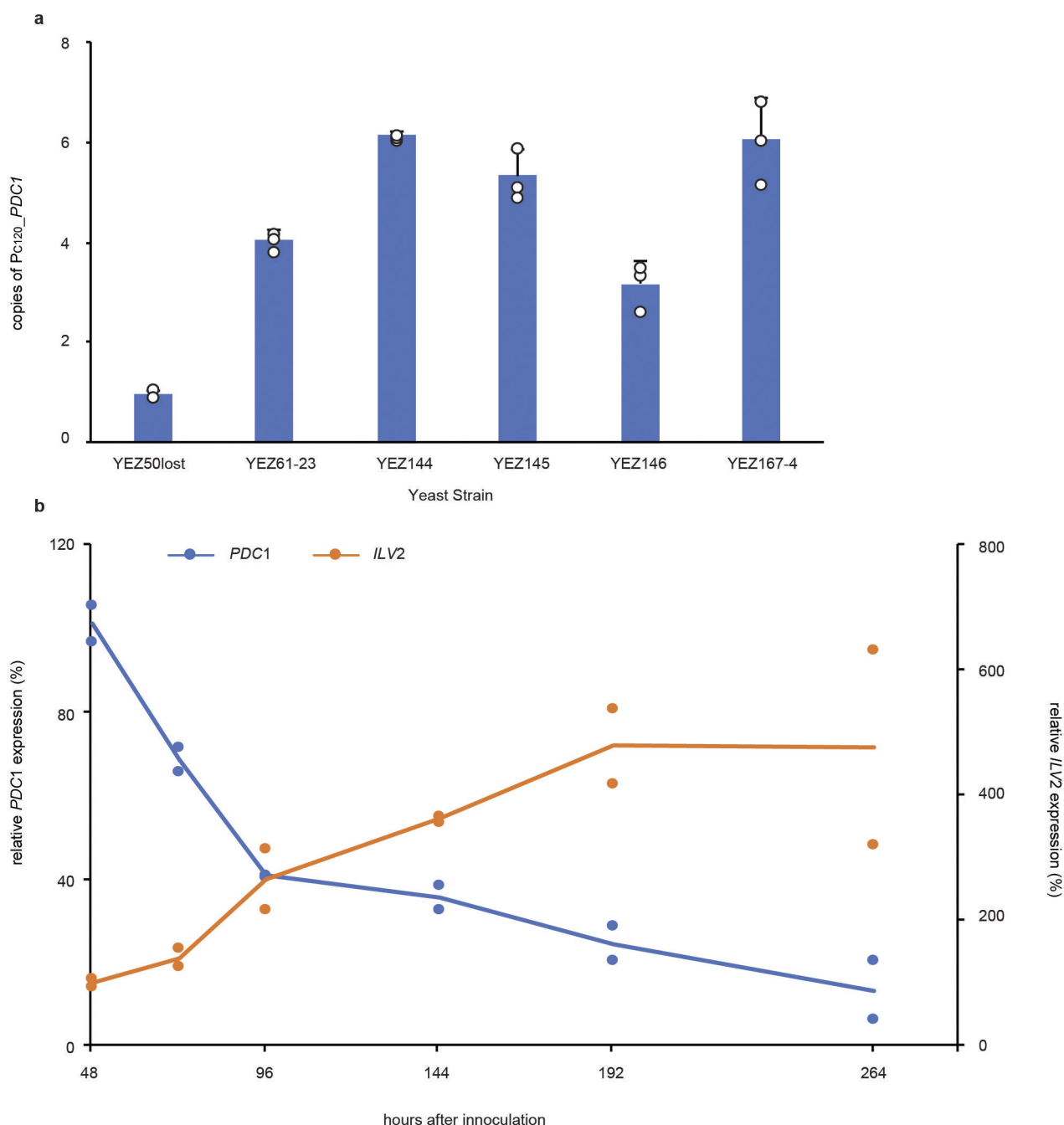
to dark and θ is the time cells are incubated in the dark before starting the fermentation ($n = 7$ biologically independent colonies). **d**, Screens for isobutanol production in 2% glucose of several colonies of YEZ159 (OptoINVRT1), YEZ156 (OptoINVRT2) and HPY6 (OptoINVRT3), using growth parameters: $\rho = 8$ and $\theta = 3$ h ($n = 12$ biologically independent colonies). **e**, Screens for isobutanol production in 2% glucose of several colonies of YEZ159 (OptoINVRT1), YEZ156 (OptoINVRT2) and HPY6 (OptoINVRT3), using growth parameters: $\rho = 5$ and $\theta = 3$ h ($n = 12$ biologically independent colonies). The screens shown in **c–e** were performed once in our laboratory.



Extended Data Figure 5 | Light-controlled lactic acid production.

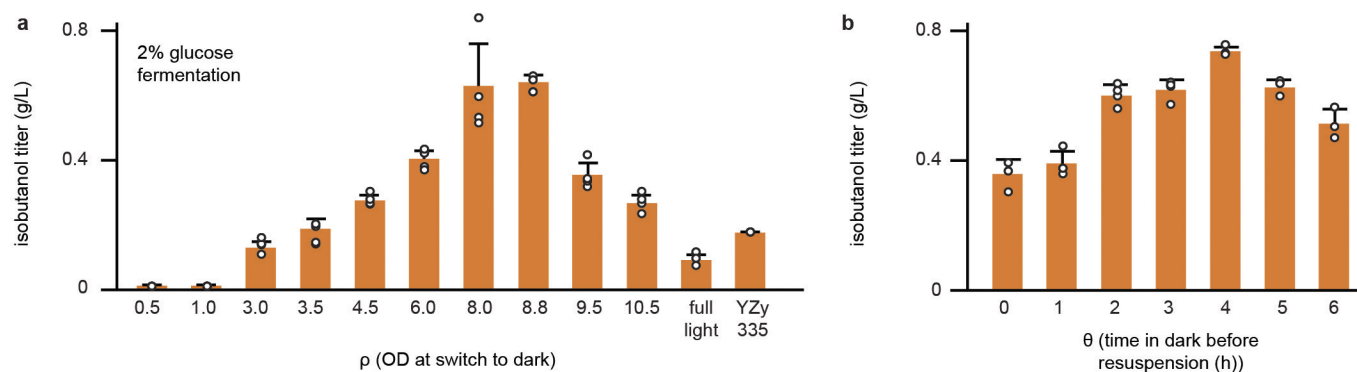
a, Lactic acid is produced by the reduction of pyruvate by Ldh. *PDC1* is controlled by OptoEXP and Ldh by OptoINVRT circuits. **b**, With optogenetic controls, light can be used to separate fermentation into two phases: a growth phase when cultures are exposed to light, during which *PDC1* is expressed and Ldh is repressed, and a lactic acid production phase when cells are in the dark, during which *PDC1* is not induced, and Ldh is expressed. **c**, Experimental design for the screening of strains and optimization of conditions. ρ and θ were varied in these experiments. **d**, Three OptoINVRT circuits were tested for lactic acid production:

OptoINVRT1 (YEZ144); OptoINVRT2 (YEZ145) and OptoINVRT3 (YEZ146). Top, dependence of lactic acid titres on ρ . Bottom, dependence of the ratio of lactic acid to ethanol on ρ . Fermentations were done in 26.5 g l^{-1} glucose and run for 2 days. All samples had $\theta = 6 \text{ h}$. Data are shown as mean values; dots represent individual data points; error bars represent the s.d. of three biologically independent 1-ml culture sample replicates exposed to the same light conditions. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Statistics are derived using a one-sided *t*-test. All experiments were repeated at least three times.



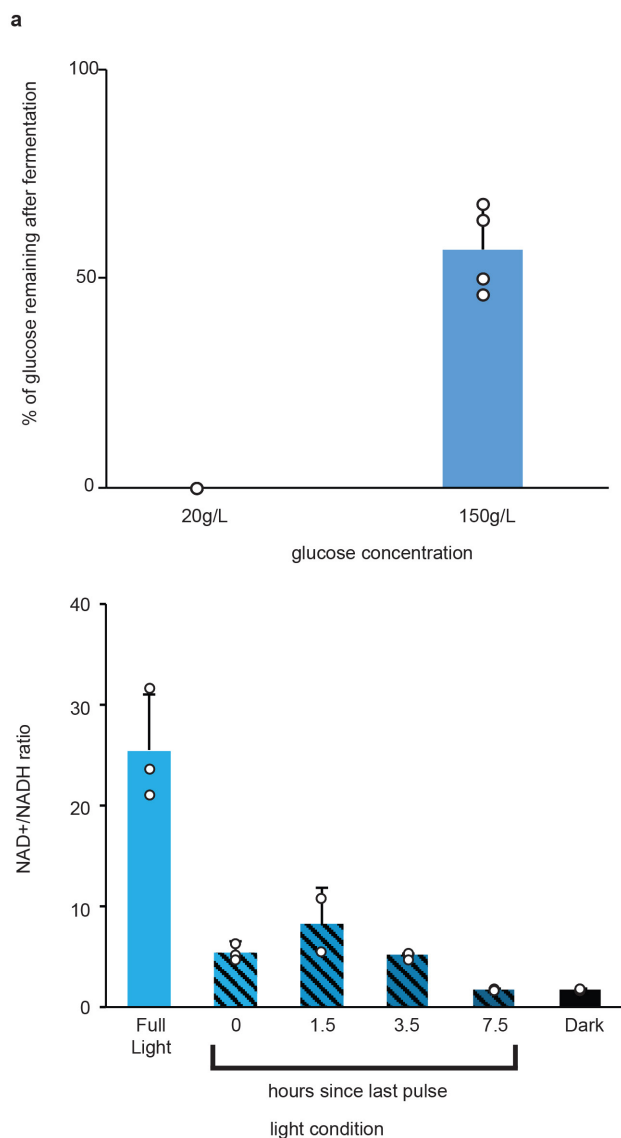
Extended Data Figure 6 | qPCR experiments. **a**, Number of copies of P_{C120} driving $PDC1$ in key strains, determined with qPCR performed on genomic DNA samples (see Methods). All strains have one copy of $PDC1$ integrated in the *HIS3* locus, and the rest are integrated in random δ -integration sites (except YEZ50lost, which only has one copy in the *HIS3* locus). Data are shown as mean values; dots represent individual data points; error bars represent the s.d. from three biologically independent 1-ml culture sample replicates. All experiments were repeated at least

three times. **b**, qPCR of $PDC1$ and $ILV2$ mRNA levels during fed-batch fermentation with periodic light stimulation for isobutanol production in 0.5-l fermenters. qPCR was performed on samples from fed-batch fermentations for isobutanol production (Fig. 3e) to measure concentrations of $PDC1$ and $ILV2$ transcripts. Gene expression was normalized to *ACT1* transcripts. Lines represent average values from samples taken from two biologically independent fermentations run under the same conditions. All experiments were repeated at least two times.



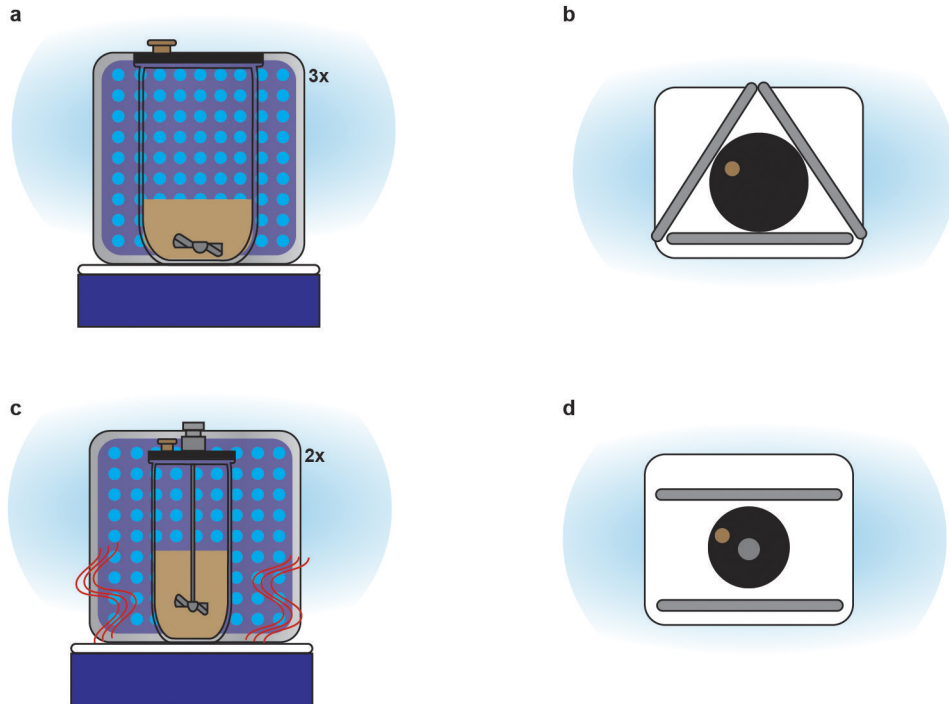
Extended Data Figure 7 | Optimization of light-controlled isobutanol production. **a**, Dependence of isobutanol titres on ρ . Cells were grown with $\theta = 3$ h; fermentations were done in 21.5 g l^{-1} glucose; isobutanol titres were measured after 2 days of fermentation in the dark. YZy335 is a control strain with a constitutive isobutanol pathway plasmid and wild-type *PDC1*, *PDC5* and *PDC6* and was used in 2-day fermentations at high

cell density as a control. **b**, Dependence of isobutanol titres on θ . Cells were grown to $\rho = 8.5$. Fermentations were again done in 20.8 g l^{-1} glucose for 2 days in the dark. All data are shown as mean values; dots represent individual data points; error bars represent the s.d. of three biologically independent 1-ml culture sample replicates exposed to the same light conditions. All experiments were repeated at least three times.



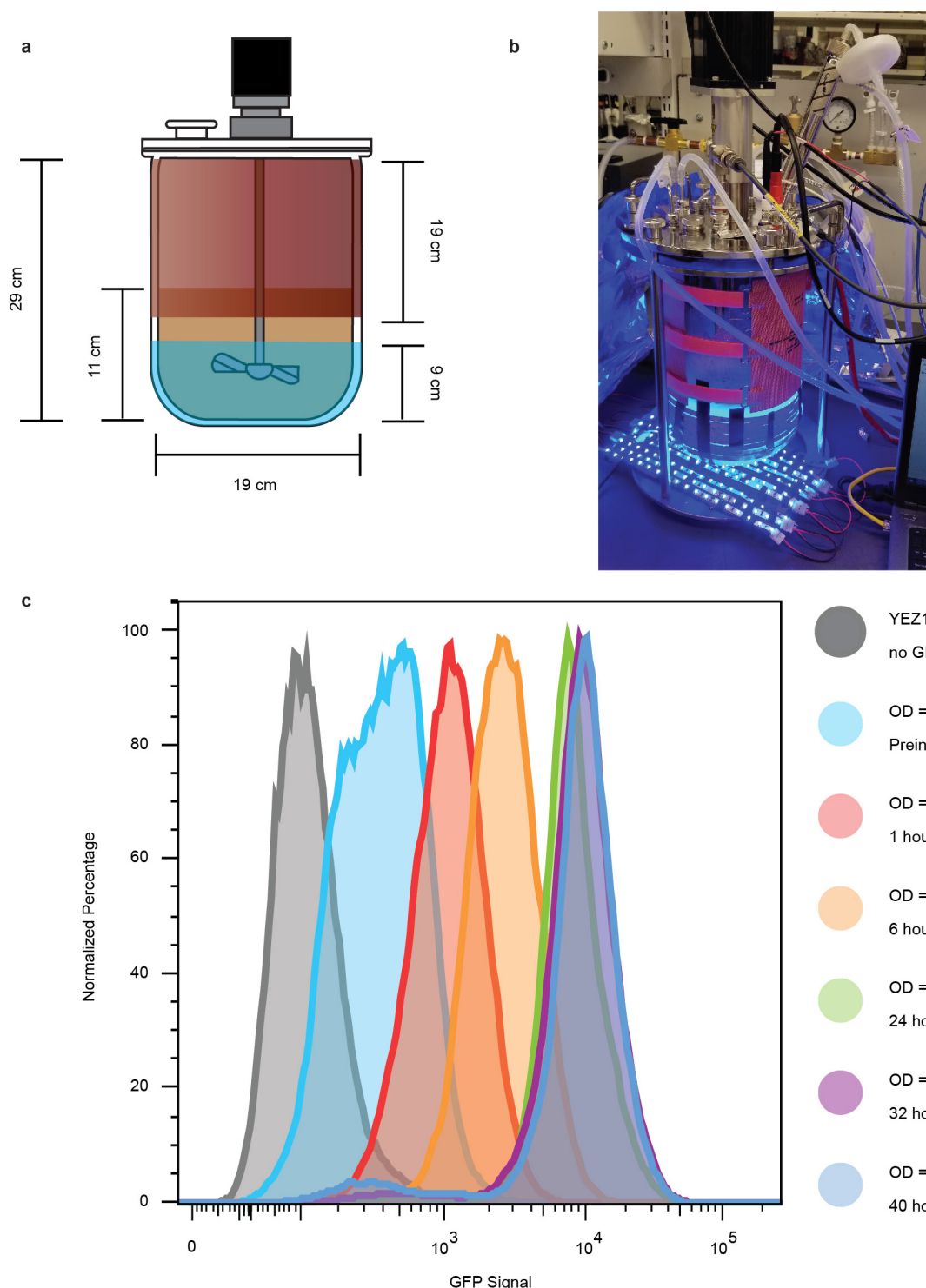
Extended Data Figure 8 | Optimization of high glucose fermentations.

a, Glucose remainders as a percentage of initial glucose concentration after 48 h (20 g l⁻¹ initial glucose) or 80 h (150 g l⁻¹ initial glucose) of fermentation of YEZ167-4 in the dark. Cell growth parameters: $\rho = 8.5$ and $\theta = 4$ h (for fermentations in 20 g l⁻¹ glucose) and $\rho = 9.5$ and $\theta = 3$ h (for fermentations in 150 g l⁻¹ glucose). Data are shown as mean values; dots represent individual data points; error bars represent the s.d. from three biologically independent 1-ml culture sample replicates. All experiments were repeated at least three times. **b**, NAD⁺/NADH ratio recovery through light pulsing. NAD⁺/NADH ratios were measured in samples under similar batch fermentation conditions as shown in Fig. 3c (see Methods). YEZ167-4 cultures were diluted into six 24-well plates and grown to an OD₆₀₀ of 9.5 and left in the dark for 3 h before resuspending cells in 15% glucose medium. Four of the plates were pulsed every 10 h for 30 min at a duty cycle of 15 s on and 65 s off. Cells were harvested after 48-h fermentations, and at different times after the last light pulse (0, 1.5, 3.5, or 7.5 h). Control plates were kept under full light or in the dark throughout the 48-h fermentations. NAD⁺/NADH ratios in pelleted cells were measured following a previously described method⁴³. Data are shown as mean values; dots represent individual data points; error bars represent the s.d. of four biologically independent 1-ml culture sample replicates. All experiments were repeated at least three times.



Extended Data Figure 9 | Diagrams of light-stimulated laboratory-scale fermenters used to test YEZ167-4. **a, b,** The 2-l bioreactor was set up so that three light panels could be placed around the fermenter. A magnetic stir plate and stir bar were used to mix the culture, and fermentations were

performed in a 30 °C warm-room. **c, d,** The 500-ml fed-batch bioreactor was set up so that two light panels could be placed around the fermenter. The culture was mixed with a motorized propeller and a heat plate with temperature control probe was used to maintain the temperature at 30 °C.



Extended Data Figure 10 | Light-dependent GFP expression in laboratory-scale fermenter at relatively high cell densities. **a**, Schematic of 5-l fermenter setup with the dimensions of the area exposed to light. Red is the heating blanket around the reactor. Brown depicts the cell culture (2.5 l). Blue depicts the area being illuminated by blue LEDs. **b**, Picture of the functioning 5-l, light-stimulated fermenter. **c**, Representative flow cytometry results from two fermentation replicates using YEZ243, which has light-inducible GFP expression. Cells were grown in fed-batch mode using a glycerol feed to achieve the highest cell densities possible in this setup. Yeast cells were exposed to light when they reached an OD_{600} of 15 and left under continuous illumination

for the rest of the experiment. Samples from the fermenter were fixed at the time of harvesting to prevent time-dependent variations. Grey was a sample of YEZ140 without GFP, which was used as a control. Light blue is pre-induction at an OD_{600} of 15; red is after 1 h of induction at an OD_{600} of 16; orange is after 6 h of induction at an OD_{600} of 19; green is after 24 h of induction at an OD_{600} of 41; purple is after 32 h of induction at an OD_{600} of 46 and dark blue is after 40 h of induction at an OD_{600} of 50. Every curve is generated from 20,000 cell counts. Data from the other fermenter run, which are very similar, are available upon request. All experiments were repeated at least three times.

ERRATUM

doi:10.1038/nature25499

Erratum: Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature

Cameron K. Ghalambor, Kim L. Hoke, Emily W. Ruell, Eva K. Fischer, David N. Reznick & Kimberly A. Hughes

Nature **525**, 372–375 (2015); doi:10.1038/nature15256

There are several errors in this Letter. First, owing to a production error, the Sequence Reads Archive (SRA) accession number is published as 'SRP06236' rather than 'SRP062364'. Second, owing to an author error, the sample size for the number of samples for the low-predation (LP) population reared without predators is listed in the Methods as ' $n = 3$ ', rather than ' $n = 4$ '. Third, owing to an author error, in Supplementary Table 1, the treatment columns referred to as 'NP' and 'P' are reversed for the HP population. Furthermore, we have identified an error in the code used to produce some of the simulations. Permutation distributions used to identify concordantly differentially expressed (CDE) genes inadvertently omitted data from the LP population. We have now reanalysed the data, including that population. In this revised analysis, all of the originally reported patterns remain unchanged, but the magnitude of statistical significance is reduced compared to the original analysis. The main results and conclusions remain unchanged, but because the reanalysis alters the set of genes deemed CDE, all subsequent tests based on the subset of CDE genes also required reanalysis. The changes to the original text, tables and figures reported in this Letter are described in more detail below.

(1) The number of adaptively evolved (CDE) genes decreased from 125 to 94, and remains outside the interquartile range for the permuted distribution (23, 73), and in the upper decile of the permuted distribution.

(2) The correlation between evolutionary change and non-adaptive plasticity (Fig. 2 of the original Letter) remains almost unchanged (-0.81 in the new analysis, and -0.82 in the original analysis).

As before, the new correlation remains well outside the interquartile range of the permuted correlations (-0.47 , -0.59), and smaller than any of the 1,000 permutation values. See Supplementary Fig. 1 to this Erratum.

(3) The correlation between ancestral and derived plasticity (Fig. 3 of this Letter) has increased from -0.588 in the original analysis to -0.743 in the new analysis. As before, the new correlation remains well outside the interquartile range of the permuted correlations (-0.57 , -0.68), and in the lower 3.5 % of the permuted values. See Supplementary Fig. 2 to this Erratum.

(4) The results comparing the observed amount of divergence in gene expression versus the amount of putative neutral divergence reported has remained almost unchanged (Extended Data Table 1 of the original Letter). The new P_{ST} values for 'Intro1' and 'Intro2' are 0.33 and 0.25 versus the original P_{ST} values of 0.32 and 0.27, respectively. All other results remain the same. Supplementary Table 1 to this Erratum contains the corrected version of the original Extended Data Table 1 that shows the comparison of gene expression divergence (P_{ST}) with divergence of putatively neutral microsatellite loci (F_{ST}).

(5) The number of significantly plastic genes increased from 565 to 658 (in Supplementary Table 2 and Extended Data Fig. 4 of the original Letter). As before, most of these transcripts exhibit plastic changes within the ancestral population that are in the opposite direction of population divergence: 67% (439 out of 658) versus 75% (424 out of 565) in the original analysis. The observation of 658 significantly plastic genes is well outside the interquartile range and in the lower 4% of permuted values. See Supplementary Fig. 3 to this Erratum.

The permutation tests used in this study are critical in determining whether the observed patterns lie outside the range of expected values given the structure of the data. The inclusion of the naturally occurring LP population in the reanalysis presented here represents a more conservative test of the occurrence of adaptively evolved CDE genes, and as a result weakens the magnitude—but not the direction—of some of the observed patterns. In summary, support for the main results and conclusions remains robust and our interpretation of the data remains unchanged. The original Letter has not been corrected online.

Supplementary Information is available in the online version of this Erratum.

CAREERS

MENTAL HEALTH Depression rates higher for grad students than general public **p.691**

GENDER Female first and last authors scarce in top journals **p.691**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

NASA



NASA scientist Josh Willis (left), who has also done stand-up comedy, celebrates success on a mission he jokingly called OMG (Oceans Melting Greenland).

LAB LIFE

How laughter in the lab can improve your science

Science is serious business — but it also shows humour can be a great thing for a team.

BY AMBER DANCE

The lab is sometimes a silly place — and perhaps it should be. A group that behaves in daft ways from time to time tends to be one that is positive, results-oriented and successful, says Michael Kerr, a business speaker in Canmore, Canada, who focuses on humour in the workplace. Jokes and pranks can serve as signs of a healthy workplace, and provide ways to foster trust and good

communication among staff, Kerr says.

A 2012 meta-review of studies on humour in the workplace found that it is linked with strong employee performance, effective stress-coping mechanisms and sturdy group cohesiveness¹. The study, conducted by two management researchers at the University of North Carolina, Wilmington, and a psychologist at Florida International University in Miami, also linked humour to reduced burn-out among employees.

Although humour has its benefits, researchers caution that jokes and laughs must be appropriate for the workplace and lab members should avoid making fun of each other in potentially or clearly hurtful ways. “Make sure you’re not harassing somebody or singling someone out,” says Sophie Scott, a cognitive neuroscientist at University College London. “Banter can be bullying.” And simply saying “I was joking” doesn’t undo the hurt, she adds.

As any comedian knows, attempts at ►

► humour can fall flat or even backfire. Adam Ruben, a molecular biologist at Sanaria, a biotechnology company in Rockville, Maryland, worries that humour could ruin a younger scientist's chances of being taken seriously. Ruben does stand-up comedy on the side, but keeps his major scientific talks mostly free of jokes.

Newcomers to a lab group should get to know their labmates and principal investigator well before they start joking around, advises Bernie Chun Nam Mak, a lecturer at Hong Kong Baptist University who studied workplace strategies, including humour, during his PhD programme in applied English linguistics at the Chinese University of Hong Kong.

BURST THE STRESS BUBBLE

There's no shortage of hilarity in the lab where James Utterback, a PhD student in physical chemistry at the University of Colorado Boulder, works. His greatest prank (so far — labmates beware) was inspired by a laser system that arrived in late 2015 for the group's studies of solar photochemistry. It came in a crate with metres and metres of bubble wrap.

While another group was in a meeting, Utterback and his accomplices coated their student office's floor, desks and printer with the bubbly sheets. They hid in the office so that they could catch their colleagues' shocked reaction. Both groups laughed and entertained themselves popping the plastic bubbles.

But that wasn't the last of it. When Utterback returned to his office after a conference in mid-2017, he discovered he'd been replaced — with a bubble-wrap mannequin, complete with a wig and clothes. "He had been named James 2," says Utterback. "He became kind of like our group mascot."

James 2 regularly rotated between the lab's student offices, surprising people who turned on the lights to find him diligently 'working' at their desks.

"Working with James [1] was seriously delightful," says Amanda Grennell, a freelance science writer in Missoula, Montana, who earned her PhD in August 2017 from the same lab. "Pranks gave my brain a much-needed break from both work and stress."

Graduate studies and science in general can be frustrating and isolating, agrees Jorge Cham of Los Angeles, California, who earned a PhD in mechanical engineering from Stanford University in California. He started a PhD comic strip, 'Piled Higher and Deeper', soon after beginning graduate school, as a sort of art therapy to cope with the pressure cooker of academic training.

After a stint as an instructor at the California Institute of Technology in Pasadena, he switched in 2005 to cartooning full-time. He

regularly lectures about the misery of graduate studies and the joys of procrastination.

This unusual career path has given Cham more fame than he probably ever would have earned in academic robotics — his original plan. The comics, which have drawn 188,000 followers on Twitter, have appeared in more than 50 US newspapers and been collated into 5 collections that have collec-



Helena González teaches science through comedy.

tively sold more than 100,000 copies. They've helped many a stressed graduate student laugh through tough times, simply by showing readers that they're not alone in their academic struggles. "Burnout is so common, anything that can be done to help people keep perspective and enjoy what they're doing, the better the long-term prospects of the lab are," says Cham.

A light attitude can also help to diffuse tension that arises from failures or errors. Cham once spilt a bucket of mildly toxic resin in another lab that had a group leader he found particularly intimidating. He was horrified. "I vividly remember my life flashing before my eyes," recalls Cham. "I never thought that was something that actually happened."

But the postdoc supervising Cham took it good-naturedly, simply telling him he was allowed to make that mistake — once. That

helped Cham to move past the error.

As that supervisor so deftly illustrated, humour can be a powerful tool for leaders. The meta-analysis of workplace humour found that when those in power were viewed as fun and funny, their subordinates performed better and had stronger teamwork. Staff were also happier with their jobs and their bosses. And a fun, lighthearted lab group might produce better science than one that is perpetually solemn and serious. Humour, says Kerr, is a positive catalyst for thinking creatively. Or, for mathematicians: "Ha + Ha = Aha."

Light-heartedness can also reduce embarrassment when inevitable mistakes happen in the lab. "Humour can help workers, especially superiors, to imply something negative to each other in a less face-threatening way," says Mak. He once observed a situation in which a worker made a mistake on an Excel spreadsheet. Her supervisor admonished her by making a wry joke about the error, and they both laughed it off.

LEAD WITH LAUGHTER

A humour-filled lab might not happen spontaneously, so leaders might wish to actively encourage it. Scott, who studies speech and laughter, makes a conscious effort to build a happy team.

But leaders should be wary of forcing humour onto junior scientists. "They have to feel comfortable first," says Scott. "They have to feel part of a group." She pays particular attention to new lab members — especially when they're from other countries — ensuring that they feel comfortable, and not left out. If they're laughing and joking, she knows that they're settling in.

To help foster a positive group dynamic, Scott often organizes lab social events and includes both personal and scientific discussions in casual lab meetings over coffee. The personal sharing is optional, of course.

"The demeanour of the principal investigator will affect the atmosphere of the whole meeting," she says. "I try to keep a positive and friendly mood."

It might seem simple for Scott's group to spice their work with humour — they study laughter, after all, and their work entails developing funny videos to make viewers and subjects chuckle. But no one has to be a stand-up comedian to infuse some humour into the lab. Kerr says that there are plenty of ways to make work fun, even for the less comedy-inclined. One option he likes is to put up a 'humour bulletin board'. Lab members can contribute funny statistics, research results or cartoons.

Starting group traditions is another way to make work enjoyable. Kerr suggests

giving out fun awards, such as the ‘Most Likely to Overdose on Caffeine Award’. Employees feel appreciated, and everyone has a good time.

Ruben recalls a fun tradition from his PhD lab at Johns Hopkins University in Baltimore, Maryland. Each scientist picked an orange test-tube stopper, or septum, and drew a little face on it. Once a week, they’d take a break to place the stoppers on a shaker; whichever stopper stayed on the longest before bouncing off won its scientist US\$1 from each of the other entrants. That two-minute break provided a valuable bonding experience, Ruben says.

And if one wonders what to joke about, “the easiest person to laugh at is yourself”, advises Josh Willis, a climate scientist at NASA’s Jet Propulsion Laboratory in Pasadena, California, who has trained with comedy troupes The Groundlings and Second City. He christened one of his missions OMG, an initialism for ‘Oceans Melting Greenland’.

Willis is perfectly willing to engage in a bit of self-deprecating humour, and one of his studies made it easy to do that. In 2006, early in his career, he collected data showing that, despite the effects of global warming, some of the world’s oceans had somehow cooled between 2003 and 2005 (ref. 2). Willis endured a fair bit of ridicule for this ‘discovery’. Even conservative pundits got in on the harassment, using Willis’s paper as evidence that left-leaning scientists are clueless about climate.

Willis, trusting his data, took it all in his stride. Then, in February 2007, he discovered the error he’d made. No, the oceans weren’t getting colder: certain temperature sensors had given bad readings. He published a correction shortly afterwards³.

In response — playing on the talk-show punditry that Willis had endured — his wife gave him a set of business cards imprinted with the title ‘Idiot leftist scientist’. He still carries the cards in his wallet. “That dose of humility and making fun of myself — in the long run, I think it benefited my career,” says Willis.

TREAD WITH CARE

Some joke topics are simply not acceptable in any workplace, notes Nicki Fuchs, a stand-up comic and biochemical engineer at MedImmune in Gaithersburg, Maryland. Gender, politics, race and religion, among other matters, are off the table, she advises.

Those rules still leave her with plenty of room to jest with her labmates. And it’s useful for her to joke about work, says Fuchs, because it helps her, a 30-year-old woman, to connect with the rest of the lab members — all older men. A recurring wisecrack is about whoever most recently left the water

running and flooded the lab — a not-uncommon occurrence, because their work involves filling up large, pressurized bottles.

Jokes may fall flat in some settings. In graduate school, Ruben often ended his lab-meeting presentations with a joke slide. But during one such presentation, his adviser was already unimpressed with his scientific progress, and Ruben suspects that his joke slide — a colleague’s head that was Photoshopped onto a goat’s body — deepened his adviser’s doubts. Since then, he uses jokes only in informal presentations or talks about science careers.

Scientists should also be careful about humour that might not work well with people from other cultures. What’s funny to one ethnic group can be incompre-

Humour is a positive catalyst for thinking creatively. Or, for mathematicians, “Ha + Ha = Aha.”

sible or offensive to another, Mak notes. It’s fairly simple to learn about the sense of humour in a geographical region to which one is travelling or moving, adds Kerr: he just Googles it.

Of course, some topics that scientists study — cancer, for example — aren’t funny at all. Yet jokes can help to ease tension and discomfort around specialities that deal with tragic subjects, says Helena González, who earned a PhD studying the epigenetics of cancer in 2013 and is now a science communicator with the comedy troupe Big Van Science in Barcelona. “That kind of black humour releases your feelings and makes your work much easier,” she says.

Still, she adds, scientists need to be careful about where they make any such jokes. Generally, among a few close colleagues in the lab, it’s fine. When dealing with patients or the public, it’s not.

And although pranks are fun, those that might endanger personnel, equipment or experiments have no place in the lab. Ruben recalls one supervisor he had at a summer job, who dropped a lit match into a recently emptied jug of ethanol. “A column of fire shot up to the ceiling,” Ruben says. “He probably shouldn’t have done that.”

That said, everyone can use a chuckle now and then. If you’re planning a — harmless! — April Fool’s lab jape, be sure to share it with @naturejobs, hashtag #AprilFools. ■

Amber Dance is a freelance writer in Los Angeles, California.

1. Mesmer-Magnus, J. *et al.* *J. Manage. Psychol.* **27**, 155–190 (2012).
2. Lyman, J. M. *et al.* *Geophys. Res. Lett.* **33**, L18604 (2006).
3. Willis, J. K. *et al.* *Geophys. Res. Lett.* **34**, L16601 (2007).

MENTAL HEALTH

Degrees and depression

PhD and master’s students worldwide report rates of depression and anxiety that are six times higher than those in the general public (T. M. Evans *et al.* *Nature Biotech.* **36**, 282–284; 2018). The report, based on the responses of 2,279 students in 26 nations, found that more than 40% of respondents had anxiety scores in the moderate to severe range, and that nearly 40% showed signs of moderate to severe depression. The high rates suggested by this study are alarming, says Teresa Evans, a neuroscientist at the University of Texas Health Science Center at San Antonio and the study’s lead author. She notes that students suffering from anxiety or depression might have been especially motivated to take the survey, which could have skewed the results. But she believes that the findings underscore the severity of the problem and the need for a response. Evans adds that universities should provide students with training to help them manage their time and cope with stress.

GENDER

Female authors scarce

Female authors are scarce in the pages of high-impact journals, according to an online analysis (Y. A. Shen *et al.* Preprint at bioRxiv <http://dx.doi.org/10.1101/275362>; 2018). Researchers at the University of Washington in Seattle gathered names of first and last authors from papers published between 2005 and 2017 in 15 major science and neuroscience journals, including *Nature*, *Science*, *Proceedings of the National Academy of Sciences* (PNAS), *Nature Neuroscience* and *Neuropsychology Review*. Women accounted for roughly 25% of first authors in *Nature* and *Science* and just over 35% of first authors in PNAS. Female first authors outnumbered men — by 53% to 47% — in only one journal, *Neuropsychology Review*. Women represented about 15% of senior or last-author spots in *Nature* and *Science* and just under 40% in *Neuropsychology Review*. The study found an inverse relationship between the prevalence of female authors and the journal’s impact factor: the higher the impact, the lower the proportion of women as first or last authors. Co-author Ione Fine, a neuroscientist, and colleagues suggest that journals could reduce the possibility of bias by adopting mandatory double-blind reviews. Nature Research, which includes all the Nature-branded journals, said that the brand “is committed to gender equality and our journals strive to support women in science”.

STARLESS NIGHT

Cold comfort.

BY ANDREW JOHNSTON

An ephemeral layer of crinkly ice coated the surface of the microphone, falling in glistening flakes as Matilda ran a gloved hand over the mic's surface. It was a sturdy device, but there was simply no way of knowing if it still functioned; even the most overzealous engineers didn't bother to plan for such conditions. The broadcasting array was still working, at least, or that's what she could glean from those instruments that were still functional. Matilda had always thought that computers were supposed to function best in a cold environment, but even a being of wire and steel must have its limitations.

Matilda drew the microphone to her face, pulling down her parka just enough to allow for speech. "Explorer Nozek, broadcasting from Rogue Object X-27881. Current astronomical location is unknown. Proximity to other objects is unknown. Time since last broadcast..." She pushed the microphone aside just in time to stifle a delirious laugh. "... I guess I don't know how long it's been. Funny, I hadn't really thought about it until just right now, but without a star to revolve around or a nice neat rotation, these rogue planets don't have 'time' like we do back on Earth, nice neat cycles that we can chart and predict and obsess over. Different circumstances, this might be very liberating."

Matilda buried her face in the parka as she fell prey to a violent coughing fit. The 'cold' she had contracted was clearly psychosomatic — she was in a hypochondriac's dark dream, a place so frigid that no pathogen could possibly survive in the air. Nothing could live here for long, or at least that was what Matilda — what everyone on Earth — had assumed before she landed.

"Sorry about that," said Matilda, idly tapping one of the displays. "OK. Current temperature is about -230°C and dropping fast as we continue to move away from all known stars. Guess I should switch over to kelvin soon, that's... a pleasant moment, when you realize the freezing point of water has become grossly inadequate for what you are experiencing.

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

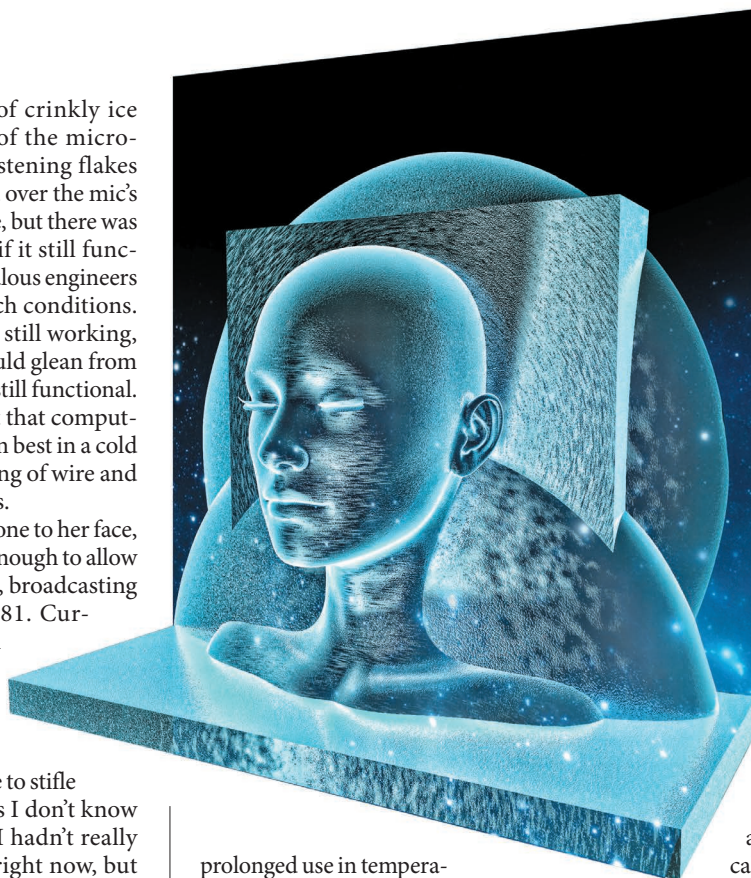
Um... the heat pump is still working, which is a minor miracle as it's really only built for

prolonged use in temperatures some 50 degrees higher than this. I expect it'll overload and break soon, probably around the time the back-up batteries give out and —"

A strange rhythmic tapping sound split the silence, a tangle of tiny feet skittering across the metal plates in the shadows at the edge of the cramped chamber. Matilda shot a quick glance out of the corner of her eye at the new arrival but didn't halt for more than a second.

"In concluding what I suspect will be my final missive, as... well, I have no clue where I am so there's no chance of anyone else finding me... anyway, I'd just like to say that the Rogue Exploration Project was a terrible idea. We're never going to know how many astronauts we've killed on this little journey, but I suspect the answer is all of us. *Oh no, sending robots isn't good enough. We need human eyes on those planets. Consider the possibilities.*' I can remember saying this myself, both while drunk and stone-cold sober, so I guess this is partially on me. If you do find me, don't bother with a funeral or service, but go ahead and dig up what's under my ship. Trust me on this."

Matilda's final words vanished into another bout of coughing, one that she



wilfully allowed the microphone to pick up — the better to let the recipients know what the end would look like. Terminating the broadcast, she let her head roll to one side, staring off into the shadows at the edge of the chamber.

"You come to watch me die?" said Matilda. "Come on out, don't skulk."

At first, she could see only the thing's deformed-looking eyes, those recessed pits that glimmered in the dim light. Then, with a series of clicks, it emerged into the shimmery light. First came the head, triangular and with few features aside from its eye-pits and a weird sucker-mouth whose mechanics were beyond knowing. Next came its slender body, jet black and covered in translucent fibrous hairs that turned rigid in time with its respiration. The odd little creature was supported by two asymmetrical lines of spindly appendages that clicked almost musically as the thing advanced.

"You alone this time, or did you bring your brothers?" Matilda sighed and slumped down in her seat, then suddenly grabbed the microphone and yanked it forward. "Come on, make a sound. So you can't speak, but can't you growl or bark or... burble or groan or something? Can't you prove that you exist before I die? Won't you at least let me be known as the one who discovered you?"

The creature merely cocked its head, moving itself up and down on its appendages. It drew closer to Matilda like a curious dog, then backed away to size her up. There was a faint sound from the bowels of the vessel, or rather the absence of a sound that Matilda had ceased to notice.

"That would be the heat pump going." Matilda glanced at the creature as the cold enveloped her. "I guess the Universe is always going to keep its secrets, huh?"

Matilda closed her eyes and drifted off to a sleep from which she would never wake. The creature's eye-pits never left her. ■

Andrew Johnston is a former Pacific-hopping ESL teacher currently scrapping away in the dark spiritual heart of Kansas.

ILLUSTRATION BY JACEY



Stockholm is home to several entrepreneurial companies that are reshaping Sweden's scientific industry.

Science diversified

After suffering heavy blows, Swedish science shows signs of sustainable recovery.

BY NIC FLEMING

Mergers, acquisitions and restructuring are par for the course in the pharmaceutical industry. Markets grow and shrink, new treatments are developed and corporate priorities change. Jobs are created, moved and lost. This came as no comfort to the 600 people left out of work when AstraZeneca shut down its research facility in Lund, Sweden, in 2011. As the ripples spread, some saw it as the latest in a series of blows signalling the death of a local industry.

Seven years on, and although Sweden's pharmaceutical sector has yet to recover, those taking a wider view see a more positive picture emerging. Major new materials-science

facilities in Lund are expected to open up prospects in a wide range of fields.

"AstraZeneca left a big black hole in a small country," says Tina Persson, a career coach in Stockholm. "My feeling is that things are growing again, but it's a new environment with lots of start-ups and innovative small companies. Those looking for work in the sector need to understand how the culture has changed so they can best find where their skills are in demand."

Sweden has a tradition of investing in research and development (R&D): it spends 3.4% of its gross domestic product on science, placing it fifth in the world. This helps to

explain why the country came second (behind South Korea) in the 2017 Bloomberg Innovation Index. Policymakers used to worry that this investment was not always matched by business and employment creation, says Åsa Lindholm Dahlstrand, who studies innovation at Lund University. However, this seems to be changing as Swedes take on board trends such as the decline of big pharma, and as the country undergoes wider cultural shifts.

"It used to be that most people in Sweden expected to find a job for life with a large multinational company, and we were relatively weak in entrepreneurship," says Lindholm Dahlstrand. "It has taken a long time but the ►

► culture is becoming more individualistic, and we've seen a growing number of innovative start-ups." Sweden's unusual legislation for intellectual-property rights is also encouraging (see 'Research rewards').

DOCTOR, DOCTOR

One area seeing rapid growth is electronic health, or e-health — a broad sector that encompasses electronic medical records, virtual health-care consultations, health IT systems and telemedicine. Sweden is a highly connected society: in 2016, 93% of inhabitants had Internet access, 79% made online purchases and 77% had smartphones or portable computers — all well above the average for the European Union.

The market for virtual doctors is expanding, with two companies — Min Doktor in Malmö and KRY in Stockholm — using voice, video and text services to cut health-care providers' costs. In March 2017, 13,000 digital medical consultations took place in Sweden, which runs a public health-care system largely funded by taxpayers. "We're seeing rapid growth in the number of digital health companies and initiatives," says Nima Jokilaakso, project manager for e-health at Swecare in Stockholm, a public-private group seeking to boost Swedish health-care and life-sciences exports.

In 2016, Sweden's government and the Swedish Association of Local Authorities and Regions adopted a target of making the country a global leader in e-health by 2025. Jokilaakso, however, says politicians have failed to match ambition with action to address e-health business concerns. "They want to demonstrate they care about the sector, but there is no funding or plan connected to the issues that are relevant to e-health companies seeking to grow." These issues include regulatory hurdles and political hostility to making profits from health care.

Competition for conventional academic and industrial research roles can be stiff, but e-health is one of the areas in which doors are opening for those willing to be flexible. Companies offering virtual-doctor and telemedicine services, for example, need specialists for their telephone advice lines — and are therefore hiring medically trained PhD graduates. "Ten years ago, PhDs weren't needed in those positions, but the world is becoming more technical," says Persson.

SCIENCE CENTRES

The job market looks different in each of Sweden's three largest cities (Stockholm, Gothenburg and Malmö). In Stockholm, Persson says, R&D roles are mostly found in small- and medium-sized companies, as well as universities. There is also a high demand for those with mathematics and physics PhDs to work as data scientists and analysts in the finance industry. Persson says that a growing number of researchers are finding work consulting for

companies, and others are transferring their skills into related roles in sales, marketing and regulation.

AstraZeneca still has an R&D centre in Mölndal, near Gothenburg, where researchers focus on cardiovascular, metabolic and respiratory disease, as well as inflammation and autoimmunity. Jobs in the Malmö-Lund area are mostly at Lund University and in small businesses and start-ups, such as those in e-health and medical technology. The small pool of jobs on offer there makes networking all the more important.

"To those who come from abroad who say it is hard to get a job, I often explain that Sweden is a small country with small, narrow networks," Persson says. "The trick is to make sure you are in them." She suggests taking part in Swedish activities, going to meet-ups and making full use of alumni, expat and Facebook groups. Persson also suggests that prospective candidates try to pick up some of the language, "even though most Swedes speak English". Although not essential for science, it does help with integrating into the local culture. "It doesn't have to be perfect, but it is an important factor when job hunting, and one often forgotten by academics."

THINKING BIG

Sweden is expecting more visiting foreign researchers in the near future, thanks to the development of two world-leading research facilities in Lund. MAX IV, which opened in 2016, is a synchrotron radiation source that, when fully operational, will be about 100 times brighter than existing facilities. Synchrotrons are particle accelerators that use magnets to steer electrons around a circular vacuum tube at close to the speed of light, causing them to radiate high-energy X-rays.

THE DIVERSIFIED AND ENTREPRENEURIAL LANDSCAPE EMERGING OFFERS GREATER RESILIENCE.

Thanks to some clever engineering from Mikael Eriksson, a Swedish-born academic and pioneer of the previous three MAX synchrotrons, MAX IV's 528-metre-circumference storage ring can push electrons closer together and create brighter X-ray pulses than its predecessors. MAX IV also includes another, smaller storage ring and a linear accelerator for different types of experiment. It is being used to study materials in greater detail than ever

PROFESSOR'S PRIVILEGE

Research rewards

One unusual aspect of academic research in Sweden is the long-standing law known as professor's privilege: unusually, university employees keep the intellectual-property rights to discoveries they make.

This encourages the creation of spin-off companies, increases patenting activity and can help universities to recruit high-quality researchers. "It has attracted some good international researchers who want to work with companies and commercialize their work," says Åsa Lindholm Dahlstrand, who studies innovation at Lund University. **N.F.**

before, shining a light on chemical reactions in batteries, the precise structure of nanoparticles and the mechanisms of new drugs.

Around 2,500 researchers per year are expected to use the facility when it becomes operational in 2019. "I think MAX IV will do for Sweden and the surrounding region what CERN has done for Switzerland," says Søren Bregenholt, vice-president at Danish pharmaceutical company Novo Nordisk, referring to Europe's particle-physics laboratory near Geneva. "It will attract a lot of researchers and develop into a point of gravity for materials and biological-materials sciences."

Also under construction in Lund is the European Spallation Source, which will open to users in 2023. The facility will include a 600-metre-long proton accelerator capable of producing neutron beams two orders of magnitude brighter than those currently available. Lund University, Chalmers University of Technology in Gothenburg and Uppsala University have been recruiting researchers who can use these world-leading tools for materials science. "We are seeing new positions being opened in areas such as magnetic materials and catalysis, because universities want to position themselves to exploit these facilities," says Tomas Lundqvist, science director at MAX IV.

Few Swedes would have chosen this rocky road, with such a major decline for their country's pharmaceutical industry over the past 20 years. Despite this, there may be long-term gains from the short-term pain: as Sweden discovered, reliance on a small number of multinationals in changing times can be dangerous. The diversified and entrepreneurial landscape currently emerging offers greater resilience to the country as a whole — and new opportunities to those ready to exploit them. ■

Nic Fleming is a freelance writer based in Bristol, UK.



MARGARETA BLOOM SANDEBÄCK

Q&A Elina Berglund

Physics to fertility

Swedish particle physicist Elina Berglund developed a contraception algorithm, based on tracking ovulation rates, while working on the Higgs boson at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. Today she lives in Stockholm, where she runs a company that markets and sells her Natural Cycles app.

What led you to develop Natural Cycles?

I wanted to give my body a break from hormonal contraception, because my now-husband Raoul and I wanted to have kids in a few years. I couldn't find any effective natural alternatives; however, I learnt that analysis of body-temperature variations can pinpoint fertility. From this, I developed my own algorithm. Both the Higgs discovery and our marriage took place in 2012. While on our honeymoon, Raoul suggested others might pay to use an app based on the algorithm. We launched Natural Cycles in 2013 and moved to Stockholm in 2014.

How does Natural Cycles work and how effective is it?

Women have long used fertility-awareness-based contraception methods, but they aren't very effective because ovulation can be difficult to predict. Natural Cycles users input daily temperature measurements. They can also add menstruation dates and hormone-test results. The algorithm returns green or red days, telling users whether they need to use protection. Our recently published large study found that, taking human error into account, 7 out of 100 women per year will get pregnant while using Natural Cycles, compared with 9 in 100 for the pill and 18 in 100 for condoms.

Is Natural Cycles just for contraception?

It can also help to reduce time to pregnancy

for women trying for a baby. When we started trying, I got pregnant first time, partly because I knew exactly when I was ovulating. Our daughter Alba is now three-and-half.

Is Sweden a good place for scientists and start-ups?

Scientists in Sweden enjoy good salaries and funding, and have a lot of research freedom. It's a modern and tech-savvy society, which is especially good for start-ups. We started off in Switzerland, and when we had investment rounds we would have to get documents printed and go to a notary's office, whereas in Sweden things are quicker and less bureaucratic, and more can be done online. Things like transportation work well and it's a good climate for innovation. It's very family-friendly because you get 16 months of parental leave and the price of day care is minimal.

What's the biggest downside?

In December it gets dark around 2.30 p.m.. It can make you feel as though you want to hide until the sun comes back out. People go out and interact with others less. On the plus side, it's lovely on the first warm day of spring when everyone is out in the sun with an ice cream. ■

BY NIC FLEMING

This interview has been edited for clarity and style.

Crossing borders

A spirit of collaboration — and an engineering icon — have together supported the emergence of a major scientific hub.

BY NIC FLEMING

Engineers may know Øresund Bridge as Europe's longest road and rail bridge. Aficionados of crime dramas will forever associate the spectacular 7.8-kilometre section between Malmö and Copenhagen with the opening credits of the popular Swedish–Danish series *The Bridge*. For scientists in the region, its opening, in 2000, underpinned the development of Medicon Valley, one of Europe's strongest life-sciences clusters.

Medicon Valley consists of the island of Zealand in eastern Denmark — home to Copenhagen, the country's capital — and the Skåne region of southern Sweden (see 'Science hub'). Its population of 4 million amounts to only around one-quarter of the combined population of the two countries. Yet the valley's 41,000-strong life-sciences workforce accounted for 58% of people working in the sector across both nations in 2015.

The hub is going through a process of change against a backdrop of global disruption, creating both challenges and opportunities for researchers nearby. Some changes are playing out differently on either side of the bridge.

The 2008–15 period, for example, saw employment in the life sciences drop by 31% in Skåne, but jump by 13% in the Danish part of Medicon Valley.

This can largely be explained by events in the pharmaceutical industry. Most prominently, multinational giant AstraZeneca shut its research facility in Lund, with the loss of 600 jobs, in 2011. Growth on the Danish side has been driven by

Novo Nordisk in Bagsværd, which has around 17,500 employees in the region, as well as by smaller pharmaceutical companies.

MAKING MEDICINE

For early-career researchers intending to work in the region's pharmaceutical sector, demand for solid computing knowledge has grown with the rise of big data — many more businesses now need trained number crunchers. "Once, statisticians could primarily get work in the pharmaceutical industry, insurance and finance, but now they are in demand all over, so we face more competition to recruit them," says Erik Kristensen, a human-resources executive at Novo Nordisk.

The company — which produces drugs to manage diabetes, obesity and haemophilia — is also looking for epidemiologists, pharmacologists, protein scientists, medical doctors and biologists focused on metabolic disease.

However, although Medicon Valley's pharmaceutical companies are hiring for specific positions, hundreds of jobs have been lost in the sector over the past three years. The region's companies are not immune to global pressures facing the industry, not least the long-term decline in research and development productivity. Big pharma's issues have, however, opened up opportunities for others. The biotechnology industry took off in the region in the 1990s, before the bridge was built. Today, there are around 150 biotech companies in the region. Europe's biggest biotech, Genmab, has

its headquarters and a clinical development team in Copenhagen. Founded in 1999, it now employs some 250 people and has 2 approved immunotherapeutic-antibody treatments. Genmab is also working on other antibody therapies, and has more than 100 ongoing clinical trials for various therapies.

For Jan van de Winkel, Genmab's chief executive, Medicon Valley is ideally placed thanks to the proximity of major Danish pharmaceutical players, and Swedish universities and hospitals across the bridge. "Being networked and gaining inspiration from others is key to Genmab's success," says van de Winkel. "I see a very positive future for Medicon Valley as the connections between different parts of the life-sciences ecosystem grow stronger still."

This growing emphasis on collaboration and networking is underlined by the rise of the region's science parks, where researchers, entrepreneurs and established businesses learn from each other. Before the bridge and the creation of Medicon Valley, the only science parks in the area were Ideon in Lund, Medeon in Malmö and Symbion in Copenhagen. Since 2004, three more have opened. Today, one in ten of those working in the life sciences in the region is based in a science park.

The Copenhagen Bio Science Park (COBIS) was set up in 2009 to support early-stage life-sciences organizations. It currently hosts more than 100 start-ups and employs around 350 people. COBIS rents out facilities, offers business-development services

Øresund Bridge, joining Sweden and Denmark, supports a major science hub.



and provides early-stage investment funding to promising projects.

“There’s a lot of buzz in the sector right now,” says Morten Mølgaard Jensen, COBIS chief executive. “The number of new projects and companies spinning out of institutions is on the rise. We offer an environment where everyone knows each other and start-ups can get established while enjoying cheap offices and access to shared lab space.”

CROSSING OLD BORDERS

Medicon Valley’s nine universities employ nearly 7,000 life-sciences researchers. Of these, the University of Copenhagen comes out top in terms of international rankings, publication volume and citations. As elsewhere, interactions between academia and industry have grown, thanks in part to policy initiatives such as the strengthening of technology-transfer offices, and wider cultural changes, including how academics view industry.

Nils Brünner, a cancer researcher at the University of Copenhagen, is one of many academics who have embraced this environment. He helped to launch the start-up Scandion Oncology in May 2017. The company is investigating a drug candidate that has shown promise in enhancing the effects of some cancer treatments by overcoming drug resistance.

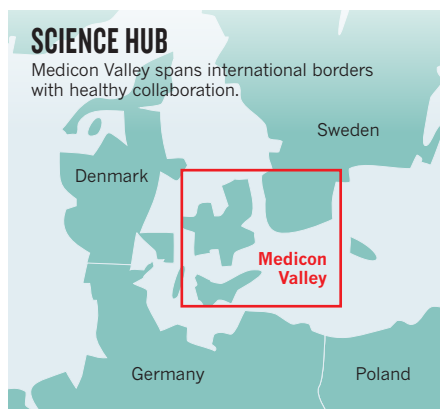
Brünner says that ten years ago, as an academic researcher, you wouldn’t want to tell people you were partnering with industry “because they might think you have a conflict of interest. That has changed very rapidly, and today academics are proud to work with businesses, and if you spin out a company you are a hero.”

Brünner, a Dane born in Sweden, set up an oncology-networking organization in 2016 with the help of the Medicon Valley Alliance (MVA), a life-sciences industry body. This seeks to drive progress against cancer through the sharing of ideas between academics and industry, from both sides of the bridge. The MVA has also set up similar medical technology and microbiome networks in the region.

INTERNATIONAL IMPACT

After the departure of AstraZeneca from Lund and several years of decline for life-sciences employment in Skåne, there are signs of recovery. The Medicon Village science park, created in 2012 at AstraZeneca’s former research facility in Lund, now hosts more than 100 businesses, including biotech companies. Also in Lund are the MAX IV synchrotron radiation facility, completed in 2016, and the European Spallation Source, an advanced pulsed neutron source currently under construction. Next-generation materials-science facilities are expected to enhance Sweden’s reputation as a leading materials-science hub, providing insights across a wide range of fields and creating job opportunities in nearby universities.

Those who promote Swedish–Danish



collaboration acknowledge that life sciences still face the challenges in the region. “We have everything Boston has,” says Søren Bregenholt, chairman of the MVA and a corporate vice-president at Novo Nordisk. “We have top universities, great hospitals, pharmaceutical companies, biotech, a well-educated workforce, a flexible job market and competences that stretch from early research to manufacturing. However, we are more diluted geographically.” Life-sciences leaders and politicians are discussing new measures to combat this disadvantage, such as improved cross-border research and development, knowledge exchange, clinical-trial collaboration, and coordination of efforts to attract investment to the region.

Another issue is a lack of availability of local capital, which can leave small but promising companies dependent on investment from abroad and vulnerable to foreign acquisition. “We have start-ups and grown-ups, but we lack the scale-ups,” says Bregenholt. In 2015, for example, Danish oncology company EpiTherapeutics was snapped up for US\$65 million by US company Gilead in Foster City, California.

Large companies in the area also report some recruitment difficulties. “People can get higher salaries in the United States and Switzerland,” says Kristensen. “Because of things like free health care and education, good transport infrastructure and a good work–life balance, however, we find that those that come tend to stay.”

To outsiders, a life-sciences hub that crosses a national border and 8 kilometres of cold water might seem strange. Yet Skåne was part of Denmark until the seventeenth century, and the two languages are very close. Øresund Bridge has brought the neighbours closer still.

“Danes understand Swedes and Swedes understand Danes,” says Brünner. “They have more or less the same cultural background. Of course, there are differences in regulations, research focuses and viewpoints. But it’s being exposed to these that opens up possibilities to find more collaborators and achieve higher standards.” ■

Nic Fleming is a freelance writer based in Bristol, UK.